In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

3/17/65
b

SOME ASPECTS OF ERGODIC THEORY

A THESIS

Presented to

The Faculty of the Graduate Division

by

Kent Bruce Erickson

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Applied Mathematics

Georgia Institute of Technology

June, 1966

SOME ASPECTS OF ERGODIC THEORY

Approved:

Chairman

Date approved by Chairman: 4 June 66

ACKNOWLEDGMENTS

I gratefully acknowledge my debt to Dr. James W. Walker for his assistance and encouragement in the preparation of this paper. I also express my thanks to Dr. Eric R. Immel for his many valuable suggestions and to Dr. Joseph Ford for his help. My thanks also go to Mrs. Betty Sims for her excellent typing of the manuscript.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS. ii

Chapter

I. INTRODUCTION 1

II. MEAN CONVERGENCE 4

III. POINTWISE CONVERGENCE. 16

IV. MEASURE PRESERVING TRANSFORMATIONS 54

BIBLIOGRAPHY 81

CHAPTER I

INTRODUCTION

Ergodic theorems are concerned with convergence of averages of iterations of an operator acting on a function space or more generally on a topological linear space.

The first result of ergodic theory was proved by J. von Neumann about 1930 and published in 1932. The von Neumann mean ergodic theorem states that if T is a measure preserving transformation on a measure space $(X, \mathfrak{A}, \mu)$, then to every $f \in L_2(X, \quad, \mu)$ there is function $f^* \in L_2$ such that

$$\lim_n \int \left| f^*(x) - 1/n \sum_{k=0}^{n-1} f(T^k x) \right|^2 d\mu = 0 \ .$$

At about the same time G. D. Birkhoff proved under additional restrictions on the transformation T and the space X that for $f \in L_1$ the sequence $1/n \sum_{k=0}^{n-1} f(T^k x)$ is pointwise convergent to $f^*(x)$ for almost all x. The supplementary restrictions on T and X were shown to be superfluous by Khinchin in 1933, and in 1945 F. Riesz gave an ingenious proof of the theorem which now goes by the name, the Birkhoff pointwise ergodic theorem.

The transformation T induces a linear operator V_T, defined by

$$(V_T f)(x) = f(Tx) \ ,$$

which is an isometry on each L_p space $1 \leq p \leq \infty$, i.e. $\|V_T f\|_p = \|f\|_p$, for every $f \varepsilon L_p$. It was such norm properties that lead S. Kakutani, K. Yosida, F. Riesz and others in the period 1935-1945 to prove various generalizations of the mean ergodic theorem to assertions concerning the convergence of operator averages in an abstract Banach space.

The first important operator-theoretic treatment of the Birkhoff theorem was given by E. Hopf in 1954. Further generalizations of the Birkhoff theorem followed, one by N. Dunford and J. T. Schwartz in 1956, another by R. V. Chacon and D. S. Ornstein in 1960. Recently (1963) Chacon has published a very inclusive generalization.

The various generalizations of the von Neumann and Birkhoff theorems take for the basic object of study a linear operator which satisfies various norm conditions. It is not assumed that the operator is induced by an underlying measure preserving transformation. The usefulness of this operator-theoretic treatment of the ergodic theorems is apparent in the study of Markov processes in which the operators are defined directly on function spaces via integral equations. Needless to say the operator V_T induced by a measure preserving transformation satisfies the hypothesis in the various generalizations so that both the von Neumann and Birkhoff theorem become corollaries.

In Chapter II of the present paper, after developing the necessary machinery from functional analysis, we proceed to a proof of a version of the Kakutani-Yosida mean ergodic theorem in a Banach space (of which the von Neumann theorem is a special case). This theorem, of interest in its own right, is used in Chapter III.

Chapter III is devoted to a development of the Dunford-Schwartz pointwise ergodic theorem along lines similar to the methods of Chacon which are based on the exploitation of properties of truncated functions.

In Chapter IV the main properties of measure preserving transformations are sketched, and the Birkhoff theorem is obtained as a corollary to the Dunford-Schwartz theorem of Chapter III. Certain properties of ergodicity are explored, and the chapter concludes with an application of ergodic theory to the theory of simple continued fractions.

Finally, we remark that this paper by no means exhausts what is by now the vast subject of ergodic theory. Thus, for example, no attempt has been made to include a discussion of application in differential equations (the original motivation of Birkhoff) or in the theory of Markov processes or weakly stationary processes. No mention has been made of the application of ergodic theory to information theory. For an introduction to these and other applications see YOSIDA, BILLINGSLEY, HALMOS [2], and DUNFORD AND SCHWARTZ [2]. For a summary of recent developments in ergodic theory see HALMOS [3].

CHAPTER II

MEAN CONVERGENCE

In this chapter we shall prove a version of the Kakutani-Yosida mean ergodic theorem. This theorem is concerned with the strong convergence of operator averages in a reflexive Banach space and will be used in the proof of the Dunford-Schwartz pointwise ergodic theorem in Chapter III.

While a general familiarity with concepts from functional analysis is tacitly assumed, this chapter is essentially self-contained. Indeed, a secondary purpose here is to give a coherent presentation of the mean ergodic theorem itself, incorporating a development of those aspects and only those aspects of functional analysis which are requisite to the proof of the theorem. Unlike the various proofs indicated in the literature, the one given here is motivated by a desire to illumine the logical development and make it reasonably complete. In this sense the organization of the proof appears to be original.

Let E be a real or complex normed linear space with norm $\|\cdot\|$. If a scalar valued function L defined on E satisfies

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y)$$

for all vectors x and y and all scalars α and β, then L is called a *linear functional* on E. The *norm* of a linear functional L is defined by

$$\|L\| = \sup\{|L(x)| : \|x\| \leq 1\} .$$

As a consequence of this definition we have:

$$|L(x)| \leq \|L\| \|x\| \quad \text{for every } x \in E .$$

It is easy to see that a linear functional L on E is continuous if and only if $\|L\| < \infty$, and it can be shown (cf. HEWITT, p. 211) that the space E^* of all continuous linear functionals on E with the usual pointwise definitions of the linear operations, and with the norm defined as above, is a Banach space (i.e., a complete normed linear space). E^* is called the *dual* or *conjugate space of* E.

LEMMA 1. *Let* M *be a linear manifold contained in E (i.e.,* M *is closed with respect to the linear operation). Let* L_0 *be a linear functional defined on* M *and suppose there exists a constant* $c \geq 0$ *such that*

$$|L_0(m)| < c\|m\| \quad \textit{for every } m \in M ,$$

then there is a linear functional L *defined on all of* E *such that*

$$L(m) = L_0(m) \quad \textit{for every } m \in M ,$$

and

$$|L(x)| \leq c\|x\| \quad \textit{for every } x \in E .$$

(Note: This last inequality implies $L \in E^*$ and $\|L\| \leq c$.)

Proof. This is an easy consequence of the Hahn-Banach theorem; cf. TAYLOR, pp. 72 and 144, or HEWITT, p. 212. ∎

LEMMA 2. *Let* $x \in E$. *If* $L(x) = 0$ *for every* $L \in E^*$, *then* $x = 0$.

Proof. Let $M = \{\alpha x: \alpha \text{ is a scalar}\}$; then M is a linear manifold in E, and the function L_o on M, defined by

$$L_o(\alpha x) = \alpha \|x\| \quad \text{for every scalar } \alpha \ ,$$

is a linear functional on M which satisfies the condition of LEMMA 1 with $c = 1$. Let L be the extension of L_o given by LEMMA 1. Since $L \in E^*$, $L(x) = 0$ by hypothesis. But $x \in M$ implies that $L_o(x) = L(x)$, and thus

$$\|x\| = L_o(x) = L(x) = 0 \ . \quad \square$$

A sequence $\{x_n\}$ in E is *weakly convergent* to a vector x, and we write $x_n \xrightarrow{w} x$, if and only if $\lim_n L(x_n) = L(x)$ for every $L \in E^*$. In view of the previous lemma it follows that a weak limit is unique. It is clear that a norm convergent sequence is weakly convergent to the same limit. The converse need not be true in infinite dimensional spaces.

If $M \subset E$ we shall denote by $\bar{M}$ the closure of M in E with respect to the norm topology in E. Thus $x \in \bar{M}$ if and only if there exists a sequence $\{x_n\}$ in M such that $\|x - x_n\| \longrightarrow 0$.

LEMMA 3. *Let* M *be a linear manifold in* E *and let* $\{x_n\}$ *be a sequence in* M *which is weakly convergent to a vector* y*; then* $y \in \bar{M}$. *Thus weak closure is equivalent to norm closure for linear manifolds.*

Proof. Let $\rho = \inf \{\|x - y\|: x \in \bar{M}\}$; it suffices to show that $\rho = 0$. Let $M_o = \{x + \alpha y: x \in \bar{M}, \alpha \text{ a scalar}\}$; then M_o is a linear manifold in

E, and the function L_o on M_o, defined by

$$(1) \qquad L_o(\alpha y + x) = \alpha\rho, \ \alpha \text{ a scalar, } x \in \bar{M} ,$$

is a linear functional on M_o. By definition of ρ we have

$$\rho \leq \left\|\left(-\frac{1}{\alpha}x\right) - y\right\| = \frac{1}{|\alpha|}\,\|x + \alpha y\|$$

for all $x \in \bar{M}$ and all scalars $\alpha \neq 0$, and hence:

$$(2) \qquad |L_o(x + \alpha y)| = |\alpha|\rho \leq \|x + \alpha y\|$$

for any $x \in \bar{M}$ and all scalars α (including $\alpha = 0$). It follows then by LEMMA 1 that there is a linear functional L on E such that

$$(3) \qquad L(z) = L_o(z) \ , \quad \text{for every } z \in M_o \ , \quad \text{and}$$

$$(4) \qquad |L(x)| \leq \|x\| \ , \quad \text{for every } x \in E \ .$$

Since $L \in E^*$ by (4), since $x_n \xrightarrow{w} y$, since $x_n \in M \subset \bar{M} \subset M_o$ and since L_o vanishes on $\bar{M}$ by (1), we have upon application of (3)

$$0 = L_o(x_n) = L(x_n) \longrightarrow L(y) \ ,$$

which implies that $\rho = L_o(y) = L(y) = 0$. ▯

The dual $(E^*)^* = E^{**}$ of the dual E^* of E is called the *second conjugate space of* E. For each $x \in E$ let $\hat{x}$ be defined on E^* by

$$\hat{x}(L) = (x), \quad \text{for every } L \in E^* \ ;$$

then $\hat{x}$ is a continuous linear functional on E^* (i.e., $\hat{x} \in E^{**}$); and, by using a Hahn-Banach type argument, it can be shown that

$$\|\hat{x}\| = \|x\| .$$

The mapping J: $x \longrightarrow \hat{x}$ is linear on E and is thus a linear isometry (i.e., norm-preserving transformation) mapping E into E^{**}. J is called the *natural* or *canonical map*.

If it happens that J actually maps E onto E^{**}, then E is said to be *norm-reflexive* or simply *reflexive*.

For a more complete discussion of the canonical map, see TAYLOR, pp. 191-192 or HEWITT, pp. 214-215.

The following lemma is very important:

LEMMA 4. *Every bounded sequence of vectors in a reflexive Banach space* E *contains a weakly convergent subsequence.*

Proof. cf. TAYLOR, p. 209. ▯

In what follows E is an arbitrary normed linear space.

A continuous linear transformation on E with range in E will be called a linear *operator* on E. The product of two linear operators S and T is defined by $(ST)(x) = S(T(x))$ for all $x \in E$. A linear transformation T on E into E is continuous if and only if it is *bounded*, i.e., if and only if $\sup \{\|Tx\| : \|x\| \leq 1\} < \infty$.

The *norm* of an operator T is the quantity

$$\|T\| = \sup \{\|Tx\| : \|x\| \leq 1\} .$$

The norm satisfies all the usual properties of a norm (provided we define addition of operators and scalar multiplication in the usual manner, viz., pointwise), and in addition we have for any operators S and T

$$\|ST\| \leq \|S\| \cdot \|T\| .$$

For any linear transformation T on E into a vector space F we shall let R[T] denote the *range* of T and N[T] the *null space* or *kernel* of T, where

$$R[T] = \{Tx : x \in E\} \subset F ,$$

$$N[T] = \{ x : Tx = 0\} \subset E .$$

It is clear that R[T] and N[T] are linear manifolds in F and E, respectively. If T is a linear operator on E, then N[T] is closed as is easily shown; R[T], however, need not be closed.

LEMMA 5. *Let* T *be a linear operator on* E*, and suppose* $x_n \xrightarrow{w} y$; *then* $T x_n \xrightarrow{w} T y$.

Proof. Let L be an arbitrary continuous linear functional on E. Define L_T on E by $L_T(x) = L(Tx)$ for all $x \in E$. Then $\|L_T(x)\| \leq \|L\| \cdot \|Tx\| \leq \|L\| \cdot \|T\| \cdot \|x\|$, so that $L_T \in E^*$. Thus, since $x_n \xrightarrow{w} y$, we have

$$\lim_n L(Tx_n) = \lim_n L_T(x_n) = L_T(y) = L(Ty) \quad .$$

Since $L \in E^*$ is arbitrary, the last equation implies that $Tx_n \xrightarrow{w} Ty$. ▯

A sequence $\{A_n\}$ of linear operators on E is said to be *strongly convergent* if and only if the sequence $\{A_n x\}$ is norm convergent for every $x \in E$. If we write $Px = \lim_n A_n y$, when $x \in E$, then P is clearly a linear transformation on E with range in E. This limiting transformation is unique, and we may rephrase the definition of strong convergence as follows:

$$A_n \xrightarrow{s} P \text{ if and only if } \|A_n x - Px\| \longrightarrow 0 \quad \text{for every } x \in E .$$

LEMMA 6. *Let* E *be a Banach space, and let* A_n *be a sequence of linear operators on* E. *Suppose that* $A_n \xrightarrow{s} P$. *Then* P *is continuous on* E *and hence is a linear operator.*

Proof. This lemma is just a slight rewording of the so-called Banach-Steinhaus theorem (cf. HEWITT, p. 218). ▯

Let M and N be linear manifolds in E. E is said to be the *direct sum* of M and N provided $M \cap N = \{0\}$ and $E = M + N = \{m + n : m \in m, n \in N\}$. We write $E = M \oplus N$.

If $E = M \oplus N$ and if $x \in E$, then there are unique vectors $m \in M$ and $n \in N$ so that $x = m + n$. It follows that the equation

$$Px = m$$

defines a linear transformation P on E with range M. This transformation is called the *projection of* E *on* M *along* N. It can be shown that a linear transformation P is a projection on some manifold M if and only if $P^2 = P$.

We are now ready for the main result of this chapter:

KAKUTANI-YOSIDA MEAN ERGODIC THEOREM (1941). *Let* E *be a real or complex reflexive Banach space, and let* V *be a linear operator on* E.
If $\|Vx\| \leq \|x\|$ *for every* $x \in E$, *then*

(i) $$E = N[I - V] \oplus \overline{R[I - V]}$$

and

(ii) $$1/n \sum_{k=0}^{n-1} V^k \xrightarrow{s} P \ .$$

where P *is the projection on* N[I - V] *along* R[I - V]. *Moreover,* P *is continuous.*

(Note: N[I - V] = {x: Vx = x} is sometimes referred to as the set of *fixed points* of V.)

Proof. Let $A_n = 1/n \sum_{k=0}^{n-1} V^k$. Since $\|V_x\| \le \|x\|$ for every $x \in E$ it follows that $\|V\| \le 1$ and hence

(1) $$\|V^k\| \le 1, \quad k = 1, 2, \ldots ; \quad \|A_n\| \le 1, \quad n = 1, 2, \ldots$$

To prove (i), let x be a fixed but arbitrary vector in E. In view of (1) the sequence $\{A_n x\}$ is bounded, and applying LEMMA 4 it follows that there is a vector y and a subsequence $\{A_{n_k} x\}$ such that

(2) $$A_{n_k} x \xrightarrow{w} y \ .$$

By LEMMA 5, the identity

(3) $$1/n \ (I - V^n) = (I - V) \ A_n \ ,$$

the inequalities (1) and the definition of weak convergence it follows that

$$\begin{aligned} |L[(I - V)y]| &= |\lim_k L[1/n_k(I - V^{n_k})x]| \le \\ &\le \lim_k 1/n_k \ \|L\| \cdot \|(I - V^{n_k})x\| \le \\ &\le \lim_k 1/n_k \ 2\|L\| \cdot \|x\| = 0 \end{aligned}$$

for every $L \varepsilon E^*$. Hence by LEMMA 1 $(I - V)y = 0$ or,

$$(4) \qquad y \; \varepsilon \; N[I - V] \; .$$

On the other hand, if $D_n = 1/n \sum_{k=1}^{n} (n-k)V^{k-1}$, then

$$(I - V)D_n = \sum_{k=1}^{n} (V^{k-1} - V^k) - 1/n \sum_{k=1}^{n} k(V^{k-1} - V^k)$$

$$= I - V^n - 1/n \sum_{k=1}^{n} kV^{k-1} + 1/n \sum_{k=0}^{n} kV^k$$

$$= I - V^n - 1/n \sum_{k=0}^{n-1} (k+1)V^k + 1/n[nV^n + \sum_{k=0}^{n-1} kV^k]$$

$$= I - 1/n \sum_{k=0}^{n-1} [(k+1) - (k)]V^k = I - A_n \; ,$$

and therefore

$$x - A_{n_k} x = (I - V)D_{n_k} x \; \varepsilon \; R[I - V] \; .$$

Hence by LEMMA 3 and (2) above,

$$(5) \qquad x - A_{n_k} x \xrightarrow{w} x - y \; \varepsilon \; \overline{R[I - V]} \; .$$

Combining (4) and (5), we obtain

$$x = y + (x-y) \; \varepsilon \; N[I - V] + \overline{R[I - V]} \; ,$$

or,

$$(6) \qquad E = N[I - V] + \overline{R[I - V]} \; .$$

Now suppose that $x \; \varepsilon \; \overline{R[I - V]}$, then there exists a sequence $\{y_m\}$

in E such that, if $z_m = y_m - Vy_m$, then $z_m \varepsilon R[I - V]$ and

(7) $$\|z_m - x\| \longrightarrow 0 .$$

Applying (1) and (3) we have

$$\|A_n x\| \leq \|A_n x - A_n z_m\| + \|A_n (I - V)y_m\|$$

$$\leq \|A_n\| \cdot \|x - z_m\| + 1/n\|y_m - V^n y_m\|$$

$$\leq \|x - z_m\| + 2/n\|y_m\| .$$

Hence by (7)

$$\limsup_n \|A_n x\| \leq \|x - z_m\| \longrightarrow 0$$

as $m \longrightarrow \infty$, and thus

(8) $$\|A_n x\| \longrightarrow 0 \quad \text{for every } x \varepsilon \overline{R[I - V]} .$$

Finally, suppose $x \varepsilon N[I - V] \cap \overline{R[I - V]}$. Then $Vx = x$, and therefore $A_n x = x$ for each $n = 1, 2, \ldots$. Applying (8), we have

$$x = \lim_n \|A_n x\| = 0 \quad \text{which implies that } x = 0 ;$$

that is,

(9) $$N[I - V] \cap R[I - V] = \{0\}$$

and then by (6) and the definition of direct sum

$$E = N[I - V] \oplus \overline{R[I - V]} \tag{10}$$

which completes the proof of (i).

Let P be the projection of E on $N[I - V]$ along $\overline{R[I - V]}$ which exists on account of (10). Then for every $x \in E$

$$(I - P)x \in \overline{R[I - V]}$$

and

$$Px \in N[I - V] ,$$

from which it follows that $A_n Px = Px$ and on applying (8) that

$$A_n x = A_n[Px + (I - P)x] = Px + A_n[(I - P)x] \longrightarrow Px \text{ (in norm)},$$

which is equivalent to saying that

$$A_n = 1/n \sum_{k=0}^{n-1} V^k \xrightarrow{s} P .$$

This completes the proof of (ii). Continuity of P now follows from LEMMA 6. []

A few remarks might be in order:

(1) If $\|V\| < 1$, then we can get a much stronger result, viz., that

$$\|V^k\| \longrightarrow 0 .$$

This follows from the inequality, $\|V^k\| \le \|V\|^k$.

(2) There are other results which bear certain resemblances to the above theorem and which come under the same general title, "mean ergodic

theorem"; see DUNFORD AND SCHWARTZ [2], pp. 660-668 or YOSIDA, pp. 213-215. These results are all extensions of the original von Neumann mean ergodic theorem for isometries in a Hilbert space for which a short proof, due to F. Riesz, is given in HALMOS [2], p. 16. For the KAKUTANI-YOSIDA THEOREM in its original setting, in applications to Markov processes, see KAKUTANI AND YOSIDA.

(3) The KAKUTANI-YOSIDA THEOREM, and others like it, are called mean ergodic theorems since they are applied mainly to operator averages acting on a function space for which the norm is defined by an integral. Thus the assertion (ii) of the foregoing theorem becomes, for example,

$$\lim_n \int |A_n f - Pf|^r \, d\mu = 0$$

i.e., "$A_n f$ converges to Pf in the mean of order r." Note at this point that we could not conclude from the KAKUTANI-YOSIDA THEOREM that the limit $\lim (A_n f)(x)$ exists at any point x in the underlying domain of the function f. It will be the purpose of Chapter III to discuss this type of convergence.

CHAPTER III

POINTWISE CONVERGENCE

The aim of this chapter is to develop a theorem which will enable us to infer that under certain conditions a sequence of averages will converge pointwise for "almost all x." More specifically, let $(X, \mathfrak{A}, \mu)$ be a σ-finite measure space, and let V be a linear operator defined at least on the space L_1, of all μ-integrable functions. If $\|Vf\|_1 \leq \|f\|_1$ for each $f \in L_1$ and if $\|Vf\|_\infty \leq \|f\|_\infty$ for every $f \in L_1 \cap L_\infty$, then, as will be seen, for each $f \in L_1$ the sequence $\{1/n \sum_{k=0}^{n-1} V^k f(x)\}$ converges for μ—almost all $x \in X$, i.e., for all $x \in X$ except possibly those x's in a set A, depending on f, of μ-measure zero; moreover, the limit function is integrable. Actually a stronger result will be proved. We shall show that the domain of an operator V satisfying the above norm conditions can be intended to include each L_p space $1 \leq p < \infty$, and that the convergence of $\{1/n \sum_{k=0}^{n-1} V^k f\}$ occurs in the L_p-mean sense as well as pointwise whenever $f \in L_p$ and $p > 1$. Mean convergence in the L_1 sense need not occur, however, unless $\mu(X) < \infty$. A counter example will be given in the sequel.

The organization of the proof, as well as some of the details such as LEMMA 1 and the construction of the Λ space in LEMMA 2 and its use at various places particularly in LEMMA 9, appear to be original, although much is owed to the proof in CHACON [1]. In particular LEMMA 3 is based on a construction of Chacon's.

A somewhat different approach to the present theorem, wherein lattice theoretic methods are employed, appears in DUNFORD AND SCHWARTZ [2], pp. 668-684. The theorem bears the name of the authors of this treatise and first appeared in DUNFORD AND SCHWARTZ [1].

Throughout this chapter $(X, \mathfrak{A}, \mu)$ is a fixed σ-finite measure space; i.e., $\mathfrak{A}$ is a σ-algebra of subsets of the non-empty set X, and μ is a σ-finite measure on $\mathfrak{A}$. It is assumed that $\mathfrak{A}$ is complete for μ; i.e., if $B \subset A$ and $\mu(A) = 0$, then $B \in \mathfrak{A}$. For any $A \in \mathfrak{A}$ we use the notation I_A to denote the indicator of A, so that $I_A(x) = 1$ or 0 according as $x \in A$ or $x \notin A$. We shall denote by M the vector space of all complex $\mathfrak{A}$-measurable functions defined a.e. on X and by S the linear manifold in M of all complex integrable simple functions, where as usual a simple function is one whose range is a finite set. For any $f \in M$ and any set C of complex numbers we shall let $[f \in C]$ denote the set $\{x : f(x) \in C\}$.

The L_p—*norm of a function* $f \in M$ is defined by

$$\|f\|_p = \begin{cases} \left(\int |f|^p d\mu\right)^{\frac{1}{p}}, & \text{if } 1 \le p < \infty \\ \operatorname*{ess\,sup}_{x \in X} |f(x)|, & \text{if } p = \infty ; \end{cases}$$

and we let $L_p = \{f: \|f\| < \infty\}$ for $1 \le p \le \infty$. Identifying functions which are equal a.e. on X makes it possible to consider $\|\cdot\|_p$ as a norm on L_p, and L_p becomes a complete normed linear space (Banach space) for each $p \in [1,\infty]$ (cf. HEWITT, p. 192). Note that L_p actually consists of equivalence classes of functions; however, we shall often refer to a function f in L_p. In all cases it should be clear from the context whether we are referring to the function f or to the equivalence class containing f.

We now state Hölder's inequality in a form convenient for our purposes: Let f and g ε M; then for every p ε $[1,\infty]$

$$\|fg\|_1 \leq \|f\|_p \|g\|_{p'} ,$$

where $p' = \frac{p}{p-1}$ if $p \neq 1$, and $p' = \infty$ for $p = 1$. In the case $p = 1$ the inequality is clear from the definitions; for a proof in the case $p > 1$ cf. HALMOS [1], p. 175, or HEWITT, p. 190.

A set H of measurable functions is *dense in* L_p if for each $f \in L_p$ and for every $\varepsilon > 0$ there exists an $h \in H$, depending on f and ε, such that $\|f - h\|_p < \varepsilon$. Clearly this is equivalent to saying that for each $f \in L_p$ there is a sequence $\{h_n\}$ in H such that $\lim_n \|f - h_n\|_p = 0$.

It is evident from the definitions that the space S of integrable simple functions is dense in L_p for $1 \leq p < \infty$. (If $\mu(X) = \infty$ then S is not dense in L_∞, since the constant function 1 is obviously not the uniform a.e. limit of any sequence of simple functions which are integrable). See HEWITT, p. 197, or TAYLOR, p. 379, for a more complete discussion of the density of S in L_p.

Let $f \in M$ and let a be a positive real number. The *upper section* or *truncation of* f *at* a is defined by

$$f^{a+}(x) = \begin{cases} 0, & \text{if } |f(x)| \leq a \\ (|f(x)| - a)\operatorname{sgn}(f(x)), & \text{if } |f(x)| \geq a . \end{cases}$$

The *lower section of* f *at* a is defined by $f^{a-} = f - f^{a+}$. Several interesting properties of sections are evident almost immediately from the definition; e.g.,

$$|f^{a+}| = |f|^{a+} = (|f| - a)I_{[|f| \geq a]} = (|f| - a)^{+} ;$$

and $f^{a+} \longrightarrow f$ uniformly as $a \longrightarrow 0^{+}$, since $|f^{a+} - f^{b+}| \leq b - a$ whenever $0 < a < b$.

Additional results concerning L_p-spaces, density, sectioning and other topics in analysis will be introduced as needed.

LEMMA 1. *Let* H *be dense in* L_2 *(or in any* L_r, $r \geq 1$), *and let* $f \in L_p$ *for some* $p \in [1,\infty)$. *Then there exists a sequence* $\{h_k\}$ *in* H *such that*

$$\lim_k \int |f - h_k|^{a+} d\mu = 0 \quad \textit{for every} \quad a > 0 .$$

Proof. Let $\delta > 0$. Since S is dense in L_p, there is a sequence $\{s_n\}$ in S such that $\lim_n \|f - s_n\|_p = 0$. Since mean convergence of any order implies convergence in measure, it follows that there exists at least one (in fact infinitely many) integrable simple function σ among $\{s_1, s_2, \ldots\}$ which satisfies

$$\mu[|f - \sigma| > \tfrac{1}{2}\delta] < \tfrac{1}{2} \quad \text{and} \tag{1}$$

$$\|f - \sigma\|_p < \delta/2 . \tag{2}$$

Since $\sigma \in L_2$ ($\sigma \in S \subset L_r$ for every $r \geq 1$) and H is dense in L_2, it follows by an argument similar to the one above that there exists an $h \in H$ for which

$$(3) \qquad \mu[|\sigma - h| > \tfrac{1}{2}\delta] < \tfrac{1}{2} \quad \text{and}$$

$$(4) \qquad \|\sigma - h\|_2 < \delta/2 .$$

The relations,

$$[|f - h| > \delta] \subset [|f - \sigma| + |\sigma - h| > \delta] \subset$$

$$\subset [|f - \sigma| > \tfrac{1}{2}\delta] \cup [|\sigma - h| > \tfrac{1}{2}\delta] ,$$

together with (1) and (3) above imply

$$\mu[|f - h| > \delta] \leq \mu[|f - \sigma| > \tfrac{1}{2}\delta] + \mu[|\sigma - h| > \tfrac{1}{2}\delta] < 1 ,$$

and hence

$$(5) \qquad \|I_{[|f - h| > \delta]}\|_r \leq 1 \quad (1 \leq r \leq \infty)$$

(equality may occur in (5) when $r = \infty$). If we integrate the inequality $|f - h| \leq |f - \sigma| + |\sigma - h|$ over the set $A = [|f - h| > \delta]$ and apply Hölder's inequality, we obtain

$$\int_A |f - h| d\mu \leq \int_A (|f - \sigma| + |\sigma - h|) d\mu = \|(f - \sigma)I_A\|_1 + \|(\sigma - h)I_A\|_1$$

$$\leq \|f - \sigma\|_p \|I_A\|_{p'} + \|\sigma - h\|_2 \|I_A\|_2 ,$$

so that by (2), (4), and (5) we have

$$(6) \qquad \int_A |f - h| d\mu \leq \delta .$$

So far, then, we have shown that for every $\delta > 0$ there exists an

$h \varepsilon H$ which satisfies (6), where, it is recalled, $A = [|f - h| > \delta]$.

We construct the required sequence $\{h_k\}$ as follows: let δ_k be a decreasing sequence of positive numbers with $\lim_k \delta_k = 0$, e.g., take $\delta_k = \frac{1}{k}$. For each $k = 1, 2, \ldots$, let h_k be the function in H which satisfies (6) with $\delta = \delta_k$. If a is any fixed positive number, then ultimately $\delta_k < a$ so that

$$A_k^* = [|f - h_k| > a] \subset A_k = [|f - h_k| > \delta_k]$$

and hence, from (6),

$$(7) \qquad \int_{A_k^*} |f - h_k| d\mu \leq \int_{A_k} |f - h_k| d\mu \leq \delta_k$$

for all k sufficiently large. But

$$0 \leq |f - h_k|^{a+} = (|f - h_k| - a) I_{A_k^*} \leq |f - h_k| I_{A_k^*} ,$$

therefore by (7)

$$0 \leq \int |f - h_k|^{a+} d\mu \leq \int_{A_k^*} |f - h_k| d\mu < \delta_k \longrightarrow 0$$

as $k \longrightarrow \infty$. ▯

The next lemma introduces a useful and interesting vector space.

LEMMA 2. *Let* $\Lambda = \{f : f \varepsilon M \text{ and } f^{a+} \varepsilon L_1, \text{ for every } a > 0\}$, *then* Λ *is a linear space of measurable functions and* $L_p \subset \Lambda$ *for every* $p \geq 1$, $p \neq \infty$.

Proof. Let f and g be in Λ, and let $c \neq 0$ be a complex number; then $|f|^{a/2+}$ $(= |f^{a/2+}|)$, $|g|^{a/2+}$ and $|f|^{a/|c|+}$ are each integrable for every $a > 0$. This together with the relations

$$|(cf)^{a+}| = |cf|^{a+} = |c| \; |f|^{\frac{a}{|c|}+} \quad \text{and}$$

$$|(f + g)^{a+}| = |f + g|^{a+} \le (|f| + |g|)^{a+} \le |f|^{\frac{a}{2}+} + |g|^{\frac{a}{2}+}$$

(provable by direct computations) implies that $(f + g)^{a+}$ and $(cf)^{a+}$ are in L_1 for every $a > 0$, which is equivalent to saying that $f + g$ and cf are in Λ, all of which proves that Λ is a linear space.

Now let $f \in L_p$ for some $p \in [1,\infty)$, let $a > 0$ and let $A = [|f| > a]$; then $\mu(A) < \infty$ (not necessarily true if $p = \infty$) , and

$$\|I_A\|_{p'} = \begin{cases} [\mu(A)]^{1/p'}, & \text{if } p \ne 1 \\ \le 1 , & \text{if } p = 1 \end{cases}$$

whereupon $\|I_A\|_{p'} < \infty$. The inequality $|f^{a+}| \le |f| I_A$ together with Hölder's inequality implies

$$\|f^{a+}\|_1 \le \|fI_A\|_1 \le \|f\|_p \cdot \|I_A\|_{p'} < \infty .$$

Since $a > 0$ is arbitrary, it follows that $f \in \Lambda$ and hence $L_p \subset \Lambda$. $\square$

NOTE. It follows from the foregoing proof that $f \in \Lambda$ if and only if $|f| \in \Lambda$; furthermore, if $f \in \Lambda$ and if $|g| \le |f|$ a.e., then $g \in \Lambda$. The last assertion follows almost immediately from the fact that $|g| \le |f| \in \Lambda$ implies that $|g|^{a+} \le |f|^{a+} \in L_1$ for every $a > 0$. These remarks will be used in the proof of the next lemma.

LEMMA 3. *Let* V *be a linear transformation on* Λ *with range in* Λ, *and suppose* $\|Vf\|_p \le \|f\|_p$ *for* $p = 1$ *and* $p = \infty$ *for every* $f \in \Lambda$. *Then for every* $a > 0$ *and for each* $f \in \Lambda$

$$\int_{E(f,a)} (a - |f|^{a-})d\mu \le \int |f|^{a+} d\mu$$

where
$$E(f,a) = [\sup_{n\ge 1} |1/n \sum_{j=0}^{n} V^j f| > a] .$$

Proof. We construct two sequences $\{f_n\}$ and $\{h_n\}$ in Λ as follows: let $f_o = f^{a+}$, $h_o = 0$; and, if $f_o, f_1, \ldots, f_n, h_o, h_1, \ldots, h_n$ have been defined, then

(1) $$f_{n+1} = Vf_n - \mathrm{sgn}(Vf_n)\cdot\min\{|Vf_n|, a - |f^{a-}| - \sum_{k=0}^{n} |h_k|\}$$

(2) $$h_{n+1} = Vf_n - f_{n+1} = \mathrm{sgn}(Vf_n)\cdot\min\{|Vf_n|, a - |f^{a-}| - \sum_{k=0}^{n} |h_k|\} .$$

That the sequences $\{f_n\}$ and $\{h_n\}$ are in Λ can be shown by induction: clearly f_o and h_o are in Λ since $f_o = f^{a+} \in L_1 \subset \Lambda$ and $h_o \equiv 0$. Suppose it is known that f_n and h_n are in Λ. Then $Vf_n \in \Lambda$. It follows from (2), however, that $|h_{n+1}| \le |Vf_n|$; hence $h_{n+1} \in \Lambda$ (c f. NOTE following LEMMA 2), and therefore $f_{n+1} = Vf_n - h_{n+1} \in \Lambda$. Thus, f_n and h_n are in Λ for every $n = 0, 1, 2 \ldots$.

Next we establish the relations

(3) $$|f^{a-}| + \sum_{k=0}^{j} |h_k| \le a , \quad j = 0, 1, 2, \ldots, \text{ and}$$

(4) $$V^j f = V^j f^{a-} + f_j + \sum_{k=0}^{j} V^{j-k} h_k , \quad j = 0, 1, 2, \ldots$$

valid a.e. on X.

For $j = 0$ (3) becomes $|f^{a-}| + |h_o| = |f^{a-}| \leq a$ which follows directly from the definition of the lower section f^{a-} . Suppose (3) holds for some integer $j = s \geq 0$. Then $a - |f^{a-}| - \sum_{k=0}^{s} |h_k|$ is non-negative, and using (2) it follows that

$$|h_{s+1}| \leq |a - |f^{a-}| - \sum_{k=0}^{s} |h_k|| = a - |f^{a-}| - \sum_{k=0}^{s} |h_k| ,$$

or, $|f^{a-}| + \sum_{k=0}^{s+1} |h_k| \leq a$. Thus (3) is true for $j = s + 1$ and hence is true for all integers $j = 0, 1, \ldots$.

Again, suppose that (4) has been verified for some integer $j = s \geq 0$. Then

$$V^{s+1}f = V(V^s f) = V(V^s f^{a-} + f_s + \sum_{k=0}^{s} V^{s-k} h_k)$$

$$= V^{s+1} f^{a-} + Vf_s + \sum_{k=0}^{s} V^{s+1-k} h_k$$

But $Vf_s = f_{s+1} + h_{s+1}$ by (2). Therefore

$$V^{s+1} = V^{s+1} f^{a-} + f_{s+1} + h_{s+1} + \sum_{k=0}^{s} V^{s+1-k} h_k$$

$$= V^{s+1} f^{a-} + f_{s+1} + \sum_{k=0}^{s+1} V^{s+1-k} h_k ,$$

and thus (4) holds for $j = s + 1$. On the other hand for $j = 0$ (4) reduces to the assertion $f = f^{a-} + f_o + h_o = f^{a+} + f^{a-}$ which is always true by definition of sectioning. Therefore, by induction, (4) holds for all integers $j \geq 0$.

Summing from $j = 0$ to $j = n$, $n \geq 1$, in (4) and using the identity

$$\sum_{j=0}^{n} \sum_{k=0}^{j} V^{j-k} h_k = \sum_{j=0}^{n} V^j (\sum_{k=0}^{n-j} h_k)$$

(which may be proved by writing out the left-hand side and regrouping terms), we find

$$(5) \qquad \sum_{j=0}^{n} V^j f = \sum_{j=0}^{n} V^j f^{a-} + \sum_{j=0}^{n} \sum_{k=0}^{j} V^{j-k} h_k + \sum_{j=0}^{n} f_j$$

$$= \sum_{j=0}^{n} V^j (f^{a-} + \sum_{k=0}^{n-j} h_k) + \sum_{j=0}^{n} f_j$$

$$= \sum_{j=1}^{n} V^j (f^{a-} + \sum_{k=0}^{n-j} h_k) + f^{a-} + \sum_{k=0}^{n} h_k + \sum_{j=0}^{n} f_j \ .$$

The inequality in (3) implies that $\| f^{a-} - \sum_{k=0}^{n-j} h_k \|_\infty \leq a$ for $n \geq j \geq 0$; by hypothesis V and hence V^j, $j \geq 0$, cannot increase the essential supremum of any function in Λ. Therefore

$$\sum_{j=1}^{n} \| V^j (f^{a-} + \sum_{k=0}^{n-j} h_k) \|_\infty \leq \sum_{j=1}^{n} \| f^{a-} + \sum_{k=0}^{n-j} h_k \|_\infty \leq na$$

a.e. on X. It follows from this that the term $\sum_{j=1}^{n} V^j (f^{a-} + \sum_{k=0}^{n-j} h_k)$ in the last equality of (5) is bounded a.e. in absolute value by na; and we obtain, then, from (5)

$$(6) \qquad \left|\sum_{j=0}^{n} V^j f\right| \le na + |f^{a-}| + \sum_{k=0}^{n} |h_k| + \sum_{j=0}^{n} |f_j|$$

$$\le na + |f^{a-}| + \sum_{k=0}^{\infty} |h_k| + \sum_{j=0}^{\infty} |f_j|$$

a.e., $n \ge 1$. Clearly (6) holds for $n = 0$. Now if $x \in E(f,a)$, then

$$(n_x + 1)a \le \left|\sum_{j=0}^{n_x} (V^j f)(x)\right|$$

for at least one integer $n_x \ge 0$. Taking $n = n_x$ in (6), it follows that

$$(n_x + 1)a \le n_x a + |f^{a-}(x)| + \sum_{k=0}^{\infty} |h_k(x)| + \sum_{j=0}^{\infty} |f_j(x)| ,$$

and therefore

$$(7) \qquad a - |f^{a-}| \le \sum_{k=0}^{\infty} |h_k| + \sum_{j=0}^{\infty} |f_j| , \quad \text{a.e. on } E(f,a) .$$

Now for any* $x \in E(f,a)$ either $\sum_{j=0}^{\infty} |f_j(x)| = 0$ or else $f_j(x) \ne 0$ for at least one integer $j \ge 0$. In the first case it follows from (7) that

$$a - |f^{a-}(x)| \le \sum_{k=0}^{\infty} |h_k(x)| .$$

---

*All statements involving x in the next several lines are to be interpreted as true for almost all x for which the statement is defined.

Suppose, then, that $f_j(x) \neq 0$. If $j = 0$, then $f_o(x) = f^{a+}(x) \neq 0$ which implies $f^{a-}(x) = a$, or

$$a - f^{a-}(x) = 0 = |h_o(x)| \leq \sum_{k=0}^{\infty} |h_k(x)| .$$

If $j \geq 1$, then by (1) and (2) it must happen that:

$$|h_j(x)| = |\min \{|Vf_{j-1}(x)| , a - |f^{a-}(x)| - \sum_{k=0}^{j-1} |h_k(x)|\}|$$

$$= |a - |f^{a-}(x)| - \sum_{k=0}^{j-1} |h_k(x)|| .$$

By (3), however, the quantity $a - |f^{a-}(x)| - \sum_{k=0}^{j-1} |h_k(x)|$ is non-negative. Therefore

$$|h_j(x)| = a - |f^{a-}(x)| - \sum_{k=0}^{j-1} |h_k(x)| ;$$

and once again we obtain

$$a - |f^{a-}(x)| = \sum_{k=0}^{j} |h_k(x)| \leq \sum_{k=0}^{\infty} |h_k(x)| .$$

The preceding considerations imply that

$$a - |f^{a-}| \leq \sum_{k=0}^{\infty} |h_k| , \quad \text{a.e. on } E(f,a) . \tag{8}$$

(Actually equality holds in (8) as may be seen by letting $j \longrightarrow \infty$ in (3).) The fact that $h_o \equiv 0$ and $|f^{a-}| = |f|^{a-}$, together with (8), leads to the inequalities

(9) $$\int_{E(f,a)} (a - |f|^{a-})d\mu \le \int_{E(f,a)} \sum_{k=1}^{\infty} |h_k|d\mu \le \int \sum_{k=1}^{\infty} |h_k|d\mu .$$

Now for $n \ge 0$ (3) implies that $a - |f^{a-}| - \sum_{k=0}^{n} |h_k|$ is non-negative; and, since $|Vf_n|$ is always non-negative, it follows from (2) that

$$0 \le \min \{|Vf_n| , a - |f^{a-}| - \sum_{k=0}^{n} |h_k|\} = |h_{n+1}| , \quad \text{a.e.}$$

Hence, substituting in (1), we find that

$$f_{n+1} = Vf_n - \operatorname{sgn}(Vf_n)\cdot|h_{n+1}| = (|Vf_n| - |h_{n+1}|)\operatorname{sgn}(Vf_n)$$

which implies $|f_{n+1}| = |Vf_n| - |h_{n+1}|$ a.e., and therefore

$$\int|h_{n+1}|d\mu + \int|f_{n+1}|d\mu = \int|Vf_n|d\mu .$$

But, by hypothesis $\int|Vf_n|d\mu = \|Vf_n\|_1 \le \|f_n\|_1$; hence

(10) $$\int|h_{n+1}|d\mu + \int|f_{n+1}|d\mu \le \int|f_n|d\mu , \quad n = 0, 1, \ldots .$$

For $n = 0$ (10) becomes

$$\int|h_1|d\mu + \int|f_1|d\mu \le \int|f^{a+}|d\mu = \int|f|^{a+}d\mu$$

But putting $n = 1$ in (10) gives $\int|h_2|d\mu + \int|f_2|d\mu \le \int|f_1|d\mu$. Therefore

$$\int|h_1| + |h_2|d\mu + \int|f_2|d\mu \le \int|f|^{a+}d\mu .$$

Continuing in this manner, we obtain

$$\int \sum_{k=1}^{n} |h_k| d\mu \leq \int \sum_{k=1}^{n} |h_k| d\mu + \int |f_n| d\mu \leq \int |f|^{a+} d\mu$$

for every $n \geq 1$. Letting $n \longrightarrow \infty$, it follows that

$$\int \sum_{k=1}^{\infty} |h_k| d\mu \leq \int |f|^{a+} d\mu \tag{11}$$

Combining (9) and (11) gives

$$\int_{E(f,a)} (a - |f|^{a-}) d\mu \leq \int \sum_{k=1}^{\infty} |h_k| d\mu \leq \int |f|^{a+} d\mu . \quad \square$$

LEMMA 4. *Let* V *be as in* LEMMA 3. *If* $\{g_k\}$ *is a sequence in* Λ *such that*

$$\lim_k \int |g_k|^{a+} d\mu = 0 \quad \textit{for every } a > 0 ,$$

then

$$\bar{g}_k = \sup_{n \geq 1} |1/n \sum_{j=0}^{n-1} V^j g_k| \xrightarrow{\mu} 0 .$$

<u>Proof</u>. Let a be any positive number and $E_k = [\bar{g}_k > a]$. Then by LEMMA 3 we have

$$\int_{E_k} (a - |g_k|^{a-}) d\mu \leq \int |g_k|^{a+} d\mu . \tag{1}$$

If $f \geq$ on X, then a straightforward calculation shows that

$$f^{a-} - f^{\frac{1}{2} a+} \leq \frac{1}{2} a ;$$

and, by rearranging this inequality, adding a at the appropriate point and replacing f by $|g_k|$, we obtain

$$(2) \qquad \frac{1}{2} a \le (a - |g_k|^{a-}) + |g_k|^{\frac{1}{2} a+}$$

Integrating (2) over E_k and applying (1) gives

$$(3) \qquad \frac{1}{2} a\ \mu(E_k) \le \int_{E_k} (a - |g_k|^{a-}) d\mu + \int_{E_k} |g_k|^{\frac{1}{2} a+} d\mu$$

$$\le \int |g_k|^{a+} d\mu + \int |g_k|^{\frac{1}{2} a+} d\mu$$

By hypothesis both terms of the last inequality of (3) have limit zero as $k \longrightarrow \infty$. Hence $\frac{1}{2} a\ \mu(E_k) \longrightarrow 0$, or

$$(4) \qquad \lim_k \mu[\bar{g}_k = \sup_{n \ge 1} |1/n \sum_{j=0}^{n-1} V^j g_k| > a] = 0 ,$$

which is equivalent to $\bar{g}_k \xrightarrow{\mu} 0$, since $a > 0$ is arbitrary. ▯

LEMMA 5. *Let* V *be a linear operator on* L_2*, and suppose* $\|Vf\|_2 \le \|f\|_2$ *for every* $f \in L_2$*. Then the set* H *of all functions of the form*

$$f_1 + f_2 - Vf_2 ,$$

where f_1 *is a fixed point of* V *in* L_2 *and* $f_2 \in L_2 \cap L_\infty$ *, is dense in* L_2.

Proof. L_r for $1 < r < \infty$, and in particular L_2, is a norm-reflexive Banach space (cf. HEWITT, pp. 222-231), and V on L_2 satisfies the hypothesis of the KAKUTANI-YOSIDA THEOREM of Chapter II. It follows that the set of functions of the form $f_1 + g_2 - Vg_2$ where f_1 is a fixed point of V in L_2 and $g_2 \in L_2$, is dense in L_2.

Let $f \in L_2$ and $\varepsilon > 0$. Then there is a function $h' = f_1 + g_2 - Vg_2$ of the above form such that

$$(1) \qquad \|f - h'\|_2 < \frac{\varepsilon}{2} .$$

On the other hand $L_2 \cap L_\infty$ is dense in L_2 since $S \subset L_2 \cap L_\infty$ and S is dense in L_2; hence there exists a function $f_2 \in L_2 \cap L_\infty$ such that

$$(2) \qquad \|f_2 - g_2\|_2 < \frac{\varepsilon}{4} .$$

Setting $h = f_1 + f_2 - Vf_2$, it follows that $h \in H$ and $h - h' = (f_2 - g_2) - V(f_2 - g_2)$. Application of (2) yields

$$\begin{aligned} \|h - h'\|_2 &\leq \|f_2 - g_2\|_2 + \|V(f_2 - g_2)\|_2 \\ &\leq 2\|f_2 - g_2\|_2 < \frac{\varepsilon}{2} \end{aligned}$$

Therefore from (1)

$$\|f - h\|_2 \leq \|f - h'\|_2 + \|h' - h\|_2 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon .$$

Thus, for every $f \in L_2$ and every $\varepsilon > 0$ there is an $h \in H$ such that $\|f - h\|_2 < \varepsilon$, which implies H is dense in L_2. ▯

(Note: LEMMA 5 and its proof are valid when L_2 is replaced by an arbitrary L_r, $r > 1$.)

The following lemma is based on the foregoing development. It is similar to the principal theorem of this chapter, but it carries a more restrictive hypothesis.

LEMMA 6. *Let* V *be a linear transformation on* Λ *with range in* Λ, *and suppose that for each* $f \in \Lambda$, $\|Vf\|_r \leq \|f\|_r$ *for every* $r \in [1, \infty]$. *Let* $f \in L_p$ *for some* p, $1 \leq p < \infty$. *Then the limit*

$$f^* = \lim_n 1/n \sum_{k=0}^{n-1} V^k f$$

exists a.e. on X. *Moreover, the function* f^* *is in* L_p; *and, if* $p > 1$, *then convergence to* f^* *also takes place in the mean of order* p.

Proof. Let $A_n = 1/n \sum_{k=0}^{n-1} V^k$. To prove the a.e. convergence of $\{A_n f\}$ it suffices to show that

$$(1) \qquad \overline{\lim}_n \overline{\lim}_m |A_n f - A_m f| = 0$$

a.e. on X.

The restriction of V to L_2 satisfies the hypothesis of LEMMA 5. If H is the set in L_2 referred to in that Lemma, then H is dense in L_2; and it follows from LEMMA 1 that there is a sequence $\{h_k\}$ in H such that

$$(2) \qquad \lim_k \int |f - h_k|^{a+} d\mu = 0 \quad \text{for every } a > 0 \text{ .}$$

Each h_k, being in H, has the form

$$h_k = f_{1,k} + f_{2,k} - Vf_{2,k}$$

where $f_{1,k}$ is a fixed point of V in L_2 and $f_{2,k}$ is an essentially bounded function in L_2.

Since $Vf_{1,k} = f_{1,k}$, it follows that $V^j f_{1,k} = f_{1,k}$, $j = 0, 1, 2 \ldots$, and hence that

$$(3) \quad A_n h_k = f_{1,k} + 1/n \sum_{j=0}^{n-1} (V^j f_{2,k} - V^{j+1} f_{2,k}) = f_{1,k} + 1/n(f_{2,k} - V^n f_{2,k}) \text{ .}$$

Writing $g_k = f - h_k$, then $g_k \varepsilon \Lambda$ and by (3)

(4) $A_n f - A_m f = 1/n(f_{2,k} - V^n f_{2,k}) - 1/m(f_{2,k} - V^m f_{2,k}) + A_n g_k - A_m g_k$.

By hypothesis V, and hence V^j, and $j \geq 0$, cannot increase the essential supremum of any function in Λ. Hence

$$|V^j f_{2,k}| \leq \|V^j f_{2,k}\|_\infty \leq \|f_{2,k}\|_\infty$$

a.e. on X for each j and k. From this inequality and (4) it follows that

$$|A_n f - A_m f| \leq 2(1/n + 1/m)\|f_{2,k}\|_\infty + |A_n g_k| + |A_m g_k|$$

$$\leq 2(1/n + 1/m)\|f_{2,k}\|_\infty + 2\bar{g}_k \quad \text{a.e.} ,$$

where $\bar{g}_k = \sup_{n\geq 1} |A_n g_k| = \sup_{n\geq 1} |1/n \sum_{i=0}^{n-1} V^j g_k|$. Or, since $\|f_{2,k}\| < \infty$ for each k,

(5) $$\overline{\lim}_n \overline{\lim}_m |A_n f - A_m f| \leq 2\bar{g}_k \quad \text{a.e.} ,$$

for each $k = 1, 2, \ldots$.

On the other hand it follows from (2) and LEMMA 4 that

$$\bar{g}_k \xrightarrow{\mu} 0 \quad \text{as } k \longrightarrow \infty .$$

But convergence in measure of $\{\bar{g}_k\}$ to zero implies that there is a subsequence $\{\bar{g}_k'\}$ of $\{\bar{g}_k\}$ which converges a.e. on X to zero. Since the left-hand side of (5) is independent of k, it follows that

$$0 \leq \overline{\lim}_n \overline{\lim}_m |A_n f - A_m f| \leq \lim_k \bar{g}_k' = 0$$

a.e. on X. This proves (1) .

When $p > 1$, the remaining assertions of LEMMA 6 are consequences of the KAKUTANI-YOSIDA THEOREM of Chapter II, since the restriction of V to L_p satisfies the hypothesis of that theorem and since L_p, $p > 1$, is a reflexive Banach space.

It remains only to show that, if $f \varepsilon L_1$, then $f^* \varepsilon L_1$. But

$$\|A_n f\|_1 \leq 1/n \sum_{j=0}^{n-1} \|V^j f\|_1 \leq \|f\|_1 ;$$

hence, by Fatou's lemma,

$$\int |f^*| d\mu = \int \underline{\lim} \, |A_n f| d\mu \leq \underline{\lim} \int |A_n f| d\mu \leq \int |f| d\mu < \infty ,$$

which proves the assertion. (This also shows that $\|f^*\|_1 \leq \|f\|_1$.) $\square$

We shall show later that L_1 convergence of $A_n f$ to f^* need not occur unless $\mu(X) < \infty$.

Our next goal is to show that the conditions imposed on V in LEMMA 6 may be considerably relaxed. More precisely, we shall prove that if V is a linear operator on L_1, if $\|Vf\|_1 \leq \|f\|$, for every $f \varepsilon L_1$ and if $\|Vf\|_\infty \leq \|f\|_\infty$ for each $f \varepsilon L_1 \quad L_\infty$, then there is an (essentially unique) extension of V to a linear transformation $\hat{V}$ on Λ with range in Λ such that $\|\hat{V}f\|_p \leq \|f\|_p$ for every $f \varepsilon \Lambda$ and every $p \varepsilon [1,\infty]$.

The following lemma is essentially a special case of the Riesz-Thorin convexity theorem (cf. ZYGMUND, p. 95). The proof given here follows that of Zygmund.

LEMMA 7. *Let* V *be a linear operator on* L_1, *and suppose* $\|Vf\|_1 \le \|f\|_1$ *for every* $f \in L_1$ *and* $\|Vf\|_\infty \le \|f\|_\infty$ *for all* $f \in L_1 \cap L_\infty$. *Then for any integrable simple function* f,

$$\|Vf\|_p \le \|f\|_p \quad \text{for every } p \in [1,\infty] .$$

Proof. The proof depends on the following maximum modulus principle for a strip: Let Φ be bounded and continuous on the strip

$$D = \{z = s + it : a \le s \le b, -\infty < t < \infty\}$$

(a, b finite) and analytic on the interior of D. If $|\Phi(z)| \le m$ for $z = a + it$ and $z = b + it$, $-\infty < t < \infty$, then $|\Phi(z)| \le m$ for all z in D. ZYGMUND, p. 93, includes a proof of this result.

Since LEMMA 7 is already true for $p = 1$ and $p = \infty$, we shall assume p is an arbitrary but fixed number with $1 < p < \infty$.

(I) Assume that $f \in S$ and $\|f\|_p = 1$.

Using Hölder's inequality and the fact that S is dense in $L_{p'}$, (where, it is recalled, $p' = \frac{p}{p-1}$), it can be shown that

(1) $$\|h\|_p = \sup \{|\int h \cdot g d\mu| : g \in S, 0 < \|g\|_{p'} \le 1\}$$

for any $h \in L_p$. Let g be an arbitrary integrable simple function with $0 < \|g\|_{p'} \le 1$. Then by (1) it suffices to show that

$$|\int Vf \cdot g d\mu| \le 1 ,$$

since (1) would then imply $\|Vf\|_p \le 1 = \|f\|_p$.

For each z in the strip

$$D = \{z = s + it : 0 \le s \le 1, -\infty < t < \infty\}$$

define integrable simple functions F_z and G_z on X by

$$F_z = |f|^{pz}\operatorname{sgn}(f) \quad \text{and} \quad G_z = |g|^{p'(1-z)}\operatorname{sgn}(g) .$$

Since f and g are both in S, we can write

$$f = \sum_{k=1}^{n} a_k e^{i\alpha_k} I_{A_k} \quad \text{and} \quad g = \sum_{j=1}^{m} b_j e^{i\beta_j} I_{B_j} ,$$

where $A_1, \ldots, A_n$ are disjoint, $B_1, \ldots, B_m$ are disjoint; $0 < \mu(A_k) < \infty$, $0 < \mu(B_j) < \infty$, for $k = 1, \ldots, n$, $j = 1, \ldots, m$; $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m$ are real; and the numbers $a_1, \ldots, a_n$, $b_1, \ldots, b_m$ are positive. From this it follows that

$$(2) \qquad F_z = \sum_{k=1}^{n} a_k^{pz}\, e^{i\alpha_k} I_{A_k} \quad \text{and} \quad G_z = \sum_{j=1}^{m} b_j^{p'(1-z)}\, e^{i\beta_j} I_{B_j} .$$

Now, letting $\Phi(z) = \int VF_z \cdot G_z d\mu$, then by (2) and the linearity of V we have

$$\Phi(z) = \sum_{k=1}^{n} \sum_{j=1}^{m} a_k^{pz}\, b_j^{p'(1-z)}\, e^{i(\alpha_k+\beta_j)} \int VI_{A_k} \cdot I_{B_j}\, d\mu$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{m} C_{k,j}\, e^{\gamma_{k,j} z} ,$$

where $\gamma_{k,j} = p \log a_k - p' \log b_j$ is real, and $C_{k,j}$ is a complex number not depending on z. Therefore Φ is a linear combination of exponentials in z and thus is bounded and continuous on the strip D (defined above) and analytic on the interior of D.

If $z = it$, t real, then $|F_z| = ||f|^{ip_t} \operatorname{sgn}(f)| = 1$ on the set $\bigcup_{k=1}^{n} A_k$ and $|F_z| = 0$ otherwise. Hence

$$\|F_{it}\|_\infty = \operatorname*{ess\,sup}_{x \,\varepsilon\, X} |F_{it}(x)| = 1 ,$$

and applying Hölder's inequality we obtain

$$|\Phi(it)| \leq \int |VF_{it} \cdot G_{it}| d\mu \leq \|VF_{it}\|_\infty \|G_{it}\|_1 .$$

But by the hypothesis on V $\|VF_{it}\|_\infty \leq \|F_{it}\|_\infty = 1$ and therefore

(3) $$|\Phi(it)| \leq \|G_{it}\|_1 .$$

Similarly, if $z = 1 + it$, then

$$\|G_z\|_\infty = \operatorname*{ess\,sup}_{x \,\varepsilon\, X} ||g(x)|^{p'(-it)} \operatorname{sgn}(g(x))| = 1 ,$$

and again by Hölder's inequality

$$|\Phi(1 + it)| \leq \|VF_{1+it}\|_1 \|G_{i+it}\|_\infty = \|VF_{1+it}\|_1 .$$

But $\|VF_{1+it}\|_1 \leq \|F_{1+it}\|_1$, by the hypothesis on V, and therefore

(4) $$|\Phi(1+it)| \leq \|F_{1+it}\|_1 .$$

Now $|G_{it}| = ||g|^{p'(1-it)} \operatorname{sgn}(g)| = |g|^{p'}$, and $|F_{1+it}| = ||f|^{p(1+it)} \operatorname{sgn}(f)| = |f|^p$. Hence

$$\|G_{it}\|_1 = \int |G_{it}| d\mu = \int |g|^{p'} d\mu = \|g\|_{p'}^{p'} \le 1 \ , \quad \text{and}$$

$$\|F_{1+it}\| = \int |F_{1+it}| d\mu = \int |f|^p d\mu = \|f\|_p^p = 1$$

from which, with (3) and (4), it follows that

$$|\Phi(it)| \le \quad \text{and} \quad |\Phi(1 + it)| \le 1$$

for all $t \varepsilon (-\infty,\infty)$. Hence

$$|\Phi(z)| \le 1 \quad \text{for every } z \varepsilon D \ ,$$

by virtue of the maximum modulus principle stated previously. In particular

$$|\Phi(\tfrac{1}{p})| \le 1 \ .$$

But $F_{\frac{1}{p}} = |f|^{p(\frac{1}{p})} \operatorname{sgn}(f) = f$ and $G_{\frac{1}{p}} = |g|^{p'(1-\frac{1}{p})} \operatorname{sgn}(g) =$ $= |g|^{p'(\frac{1}{p'})} \operatorname{sgn}(g) = g$ from which it now follows that

$$|\int Vf \cdot g \ d\mu| = \Phi(\tfrac{1}{p}) \le 1 \ .$$

(II) Assume $f \varepsilon S$ is arbitrary.

If $\|f\|_p = 0$, then $f = 0$ a.e. and hence $Vf = 0$ a.e. Thus $\|Vf\|_p = 0 \le 0 = \|f\|_p$.

If $\|f\|_p \ne 0$, then $\|f'\|_p = 1$ where $f' = \dfrac{f}{\|f\|_p}$. Hence by (I)

$$\frac{\|Vf\|_p}{\|f\|_p} = \| V(\frac{f}{\|f\|_p})\|_p = \|Vf'\|_p \le 1$$

which implies $\|Vf\|_p \le \|f\|_p$. ▯

LEMMA 8. *Let* V *be as in* LEMMA 7. *Then for every* p, $1 < p < \infty$, *there is a unique operator* V_p *on* L_p *which satisfies*

$$V_p f = Vf \quad \textit{for every } f \in L_1 \cap L_p .$$

Moreover $\|V_p f\|_p \le \|f\|_p$ *for every* $f \in L_p$.

Proof. Let $f \in L_p$, and let $\{\sigma_n\}$ be any sequence in S for which $\|f - \sigma_n\|_p \longrightarrow 0$. The sequence $\{V\sigma_n\}$ is a Cauchy sequence in L_p since $\{\sigma_n\}$ is Cauchy in L_p and by LEMMA 7

$$\|V\sigma_n - V\sigma_m\|_p = \|V(\sigma - \sigma_m)\|_p \le \|\sigma_n - \sigma_m\|_p .$$

Therefore since L_p is complete, it follows that there is an element $g \in L_p$ such that $\|V\sigma_n - g\|_p \longrightarrow 0$. If $\{\sigma_n'\}$ is any other sequence in S with limit f in the mean of order p and if g' is the corresponding limit as found above, then

$$\|g - g'\|_p \le \|g - V\sigma_n\|_p + \|V\sigma_n - V\sigma_n'\|_p + \|V\sigma_n' - g'\|_p .$$

But $\|V\sigma_n - V\sigma_n'\|_p \le \|\sigma_n - \sigma_n'\|_p$, and therefore

$$\|g - g'\|_p \le \|g - V\sigma_n\|_p + \|\sigma_n - \sigma_n'\|_p + \|V\sigma_n' - g'\|_p \le$$

$$\le \|g - V\sigma_n\|_p + \|\sigma_n - f\|_p + \|f - \sigma_n'\|_p + \|V\sigma_n' - g'\|_p .$$

Letting $n \to \infty$, it follows that $\|g - g'\|_p = 0$, or $g = g'$ a.e. Thus g, the limit of $\{V\sigma_n\}$ in the mean of order p, is independent of the sequence $\{\sigma_n\}$ provided only that $\|\sigma_n - f\|_p \longrightarrow 0$; and it follows that the equation

$$(1) \qquad V_p f = g$$

defines a transformation V_p on L_p into L_p which is clearly linear. Furthermore, since $\|V\sigma_n\|_p \longrightarrow \|g\|_p$ $\left(\left|\|V\sigma_n\|_p - \|g\|_p\right| \le \|V\sigma_n - g\|_p\right)$ and $\|\sigma_n\|_p \longrightarrow \|f\|_p$, we have that

$$\|g\|_p = \lim \|V\sigma_n\|_p \le \lim \|\sigma_n\|_p = \|f\|_p \ ;$$

that is

$$\|V_p f\|_p \le \|f\|_p \quad \text{for every } f \in L_p \ .$$

This last inequality of course also implies that V_p is continuous on L_p.

To see that V_p is unique let U be any operator on L_p such that

$$Uh = Vh \quad \text{for every } h \in H$$

where H is dense in L_p and $H \subset L_1 \cap L_p$. For any $f \in L_p$ let $\{h_n\}$ be any sequence in H for which $\|f - h_n\|_p \longrightarrow 0$. Then $V_p h_n = Vh_n = Uh_n$, and, using continuity of both U and V_p, we have

$$\|Uf - V_p f\|_p \le \|Uf - Vh_n\|_p + \|Vh_n - V_p f\|_p =$$

$$= \|Uf - Uh_n\|_p + \|V_p h_n - V_p f\|_p \longrightarrow 0$$

as $n \to \infty$. Hence $\|Uf - V_p f\|_p = 0$ or $Uf = V_p f$ a.e. for every $f \in L_p$.

We finish the proof by showing that V_p agrees with V on $L_1 \cap L_p$. Let f be any function in $L_1 \cap L_p$. Then there is a sequence $\{\sigma_n\}$ in S such that

(2) $$|\sigma_n| \le |f| \quad \text{a.e.,} \quad n \ge 1 \text{ ,} \quad \text{and}$$

(3) $$\|\sigma_n - f\|_p \longrightarrow 0$$

(cf. HEWITT, p. 197). Condition (3) implies $\sigma_n \xrightarrow{\mu} f$, and this together with (2) and the fact that $f \in L_1$ implies by the Lebesgue dominated convergence theorem that $\|\sigma_n - f\|_1 \longrightarrow 0$. Therefore, by continuity of V on L_1 it follows that

(4) $$\|V\sigma_n - Vf\|_1 \longrightarrow 0 \text{ .}$$

On the other hand by definition of V_p (equation (1)) and (3) we have

(5) $$\|V\sigma_n - V_p f\|_p \longrightarrow 0 \text{ .}$$

But (4) implies $V\sigma_n \xrightarrow{\mu} Vf$, and (5) implies $V\sigma_n \xrightarrow{\mu} V_p f$; and thus, $Vf = V_p f$ a.e. since limits in measure are unique (cf. HALMOS [1], p. 92). ▯

The next lemma gives the extension of V to Λ promised at the end of LEMMA 6.

LEMMA 9. *Let* V *be as in* LEMMAS 7 *and* 8. *Then for all* f *in* Λ *the limit*

$$\hat{V}f = \lim_{a \to 0^+} Vf^{a+}$$

exists uniformly a.e. on X *and defines an extension of* V *to a linear transformation* $\hat{V}$ *on* Λ *with range in* Λ. *Moreover for each* $f \in \Lambda$ *and for every* $p \in [1,\infty]$ *we have*

$$\|\hat{V}f\|_p \le \|f\|_p .$$

The transformation $\hat{V}$ *is unique in two senses:* (i) *For* $1 < p < \infty$ *the restriction of* $\hat{V}$ *to* L_p *is identical with the unique linear operator* V_p *on* L_p *determined in* LEMMA 8. (ii) *If* U *is a linear transformation on* Λ *into* M *which extends* V *and if* $\|Uf\|_\infty \le \|f\|_\infty$ *for every* $f \in \Lambda$, *then* $U \equiv \hat{V}$.

Proof. Let $f \in \Lambda$; then $f^{a+} \in L_1$ for every $a > 0$, which implies Vf^{a+} is defined. Furthermore, if $0 < a < b$, then $|f^{a+} - f^{b+}| \le b - a$ a.e. so that by the hypothesis on V we have

$$|Vf^{a+} - Vf^{b+}| \le \|V(f^{a+} - f^{b+})\|_\infty \le \|f^{a+} - f^{b+}\|_\infty \le b - a$$

a.e. on X. From this inequality it follows that the limit of Vf^{a+} as $a \to 0^+$ exists uniformly a.e., and we define

$$\hat{V}f = \lim_{a\to 0^+} Vf^{a+} \quad \text{for each } f \in \Lambda. \tag{1}$$

To see that the operation $\hat{V}$ defined by (1) in linear on Λ we proceed as follows: Let f and g be arbitrary functions in Λ, and let c be a non-zero complex number. Then to prove that $\hat{V}(f + g) = \hat{V}f + \hat{V}g$ and $\hat{V}(cf) = c(\hat{V}f)$ it suffices to show that

$$\lim_{a\to 0^+} \|V(f + g)^{a+} - (Vf^{a+} + Vg^{a+})\|_\infty = 0 \quad \text{and} \tag{2}$$

$$\lim_{a\to 0^+} \|V(cf)^{a+} - cVf^{a+}\|_\infty = 0 , \tag{3}$$

respectively. Now for any $a > 0$ we have

$$(4) \quad |(f+g)^{a+} - (f^{a+} + g^{a+})| = \begin{cases} |f^{a+} + g^{a+}|, & \text{when } |f+g| \le a \\ |f + g - \bar{a} - f^{a+} - g^{a+}|, & \text{otherwise} \end{cases}$$

where $\bar{a} = a\,\mathrm{sgn}(f+g)$. But

$$|f + g - \bar{a} - f^{a+} - g^{a+}| = |f^{a-} + g^{a-} - \bar{a}| \le 3a ,$$

and by direct though tedious manipulations

$$|f^{a+} + g^{a+}| \le |f+g| + 2a \le 3a \quad \text{when} \quad |f+g| \le a .$$

Hence from (4) it follows that $|(f+g)^{a+} - (f^{a+} + g^{a+})| \le 3a$, and therefore

$$\|V(f+g)^{a+} - (Vf^{a+} + Vg^{a+})\|_\infty \le \|(f+g)^{a+} - (f^{a+} + g^{a+})\|_\infty \le 3a \longrightarrow 0$$

as $a \to 0^+$ which proves (2).

To prove (3) suppose $|c| \ge 1$; then by direct application of the definition of upper section we obtain for each $a > 0$

$$|(cf)^{a+} - c(f^{a+})| = \begin{cases} 0 , & \text{when } |f| \le \dfrac{a}{|c|} \\ |cf| - a \le a(|c| - 1), & \text{when } \dfrac{a}{|c|} \le |f| \le a \\ a(|c| - 1), & \text{when } |f| \ge a ; \end{cases}$$

whereupon $|(cf)^{a+} - c(f^{a+})| \le a(|c| - 1)$. Similarly if $0 < |c| \le 1$, then $|(cf)^{a+} - c(f^{a+})| \le a(1 - |c|)$. In either case it follows that

$$\|V(cf)^{a+} - cVf^{a+}\|_\infty \leq \|(cf)^{a+} - c(f^{a+})\|_\infty \leq a\Big||c| - 1\Big| \longrightarrow 0$$

as $a \to 0^+$.

To show that $\hat{V}$ is an extension of V let $f \in L_1$. Then we have

$$|Vf^{a+} - Vf| \leq \|V(f^{a+} - f)\|_\infty \leq \|f^{a+} - f\|_\infty = \|f^{a-}\|_\infty \leq a$$

whereupon $\lim_{a\to 0^+} Vf^{a+} = Vf$ uniformly a.e., and by (1)

(5) $$\hat{V}f = Vf \quad \text{for every } f \in L_1 .$$

So far then we have shown that $\hat{V}$ is a linear transformation on Λ which extends V.

Before showing that the range of $\hat{V}$ is in Λ it simplifies matters to establish the inequality

(6) $$\|\hat{V}f\|_p \leq \|f\|_p , \quad \text{for every } p \in [1,\infty], \quad f \in \Lambda .$$

If $f \in \Lambda$ and if for some $p \in [1,\infty]$, $\|f\|_p = \infty$, then (6) holds trivially. Assume then that for some p f is an arbitrary function in L_p so that $\|f\|_p < \infty$. We consider three cases.

(I) Assume $p = 1$. Then by (5) $\hat{V}f = Vf$, and (6) then follows from the hypothesis on V which says $\|Vf\|_1 \leq \|f\|_1$ for every $f \in L_1$.

(II) Assume $1 < p < \infty$. Then for each $a > 0$ $f^{a+} \in L_1 \cap L_p$, which implies $Vf^{a+} = V_p f^{a+}$ by LEMMA 8; and therefore

(7) $$\int |Vf^{a+}|^p d\mu = \|V_p f^{a+}\|_p^p \leq \|f^{a+}\|_p^p$$

for every $a > 0$, since V_p does not increase the L_p norm of any function in L_p. Since $\lim_{a\to 0^+} |Vf^{a+}|^p = |\hat{V}f|^p$, it follows from (7) and an application of Fatou's lemma that

$$\|\hat{V}f\|_p^p \leq \liminf_{a\to 0^+} \int |Vf^{a+}|^p d\mu \leq$$

$$\leq \liminf_{a\to 0^+} \|f^{a+}\|_p^p \leq \|f\|_p^p$$

where the last inequality follows from the fact that $|f^{a+}| \leq |f|$ for every $a > 0$. Hence $\|\hat{V}f\|_p \leq \|f\|_p$.

(III) Assume $p = \infty$. Then $f^{a+} \in L_1 \cap L_\infty$, so that $\|Vf^{a+}\|_\infty \leq \|f^{a+}\|_\infty$ for every $a > 0$ by the hypothesis on V, and it follows that

$$|\hat{V}f| = \lim_{a\to 0^+} |Vf^{a+}| \leq \liminf_{a\to 0^+} \|Vf^{a+}\|_\infty \leq$$

$$\leq \liminf_{a\to 0^+} \|f^{a+}\|_\infty \leq \|f\|_\infty$$

a.e. on X. Therefore

$$\|\hat{V}f\|_\infty = \operatorname*{ess\,sup}_{x \in X} |(\hat{V}f)(x)| \leq \|f\|_\infty .$$

This concludes the proof of (6).

To see that the range of $\hat{V}$ is in Λ consider any function f in Λ. We must show that

$$(8) \qquad |\hat{V}f|^{a+} \in L_1 \quad \text{for every } a > 0 .$$

Let $a > 0$ be arbitrary. Since Λ is a linear space and $f^{a/3+} \varepsilon L_1 \subset \Lambda$, it follows that $f^{a/3-} = f - f^{a/3+} \varepsilon \Lambda$, whereupon by the linearity of $\hat{V}$ and various properties of sectioning we have

$$(9) \qquad |\hat{V}f|^{a+} = |\hat{V}f^{\frac{a}{3}+} + Vf^{\frac{a}{3}-}|^{a+} \le$$

$$\le |\hat{V}f^{\frac{a}{3}+}|^{\frac{a}{2}+} + |\hat{V}f^{\frac{a}{3}-}|^{\frac{a}{2}+} \le |\hat{V}f^{\frac{a}{3}+}| + |\hat{V}f^{\frac{a}{3}-}|^{\frac{a}{2}+} .$$

Now by (6) $\hat{V}$ will not increase the L_∞-norm of any function in Λ, and hence

$$|\hat{V}f^{\frac{a}{3}-}| \le \|\hat{V}f^{\frac{a}{3}-}\|_\infty \le \|f^{\frac{a}{3}-}\|_\infty \le \frac{a}{3} \quad \text{a.e.}$$

Now from the definition of upper section it follows that $|\hat{V}f^{\frac{a}{3}-}|^{\frac{a}{2}+} = 0$ a.e. Substituting in (9), we obtain

$$|\hat{V}f|^{a+} \le |\hat{V}f^{\frac{a}{3}+}| + |\hat{V}f^{\frac{a}{3}-}|^{\frac{a}{2}+} = |\hat{V}f^{\frac{a}{3}+}|$$

a.e. On integrating this inequality and applying (6) again it follows that

$$\int |\hat{V}f|^{a+} d\mu \le \int |\hat{V}f^{\frac{a}{3}+}| d\mu = \|\hat{V}f^{\frac{a}{3}+}\|_1 \le \|f^{\frac{a}{3}+}\|_1 < \infty$$

which proves (8) since $a > 0$ is arbitrary.

In order to prove the uniqueness assertion (i) let $\hat{V}_p$ denote the restriction of $\hat{V}$ to L_p where $1 < p < \infty$. Then $\hat{V}_p$ is a linear operator on L_p which extends V (cf. (5) and (6) above). But by LEMMA 8 such an extension is unique.

To prove (ii) let U be a linear transformation on Λ with range in M, let $Uf = Vf$ for every $f \in L_1$ and suppose $\|Uf\|_\infty \le \|f\|_\infty$ for every $f \in \Lambda$. Then for every $a > 0$ and for any $f \in \Lambda$ (5) and (6) imply

$$\|Uf - \hat{V}f\|_\infty \le \|Uf - Vf^{a+}\|_\infty + \|Vf^{a+} - Vf\|_\infty =$$

$$= \|Uf - Uf^{a+}\|_\infty + \|\hat{V}f^{a+} - \hat{V}f\|_\infty \le$$

$$\le \|f - f^{a+}\|_\infty + \|f^{a+} - f\|_\infty \le 2a .$$

Letting $a \to 0^+$, it follows that $\|Uf - \hat{V}f\|_\infty = 0$, or $Uf = \hat{V}f$ a.e. for any $f \in \Lambda$. ▯

We are now ready for the main theorem of this chapter.

DUNFORD-SCHWARTZ POINTWISE ERGODIC THEOREM (1956). *Let* V *be a linear operator on* L_1 *with* $\|Vf\|_1 \le \|f\|_1$ *for every* $f \in L_1$ *and* $\|Vf\|_\infty \le \|f\|_\infty$ *for every* $f \in L_1 \cap L_\infty$. *Then for each* $f \in L_1$ *the limit*

$$f^* = \lim_n 1/n \sum_{k=0}^{n-1} V^k f$$

exists a.e. on X; *the limit* f^* *is integrable, in fact* $\|f^*\|_1 \le \|f\|_1$, *and is a fixed point of* V.

Proof. The proof follows from LEMMAS 6 and 9. For if $\hat{V}$ is the extension of V given by LEMMA 9, then $\hat{V}$ satisfies the hypothesis of LEMMA 6; and it follows that the limit

$$f^* = \lim_n 1/n \sum_{k=0}^{n-1} \hat{V}^k f \tag{1}$$

exists a.e. for any f in L_p, $1 \leq p < \infty$. In particular if $f \in L_1$, then we may replace $\hat{V}$ by V in (1). That

(2) $$\|f^*\|_1 \leq \|f\|_1$$

follows as in the proof of LEMMA 6.

To see that f^* is a fixed point of V let $\{g_k\}$ be a sequence of integrable simple functions such that

(3) $$\lim \|f - g_k\|_1 = 0 .$$

Since $\{g_k\}$ is in $L_1 \cap L_p$ for every $p > 1$, it follows by LEMMA 6 that, taking $p = 2$,

(4) $$\lim \|A_n g_k - g_k^*\|_2 = 0 \quad (k \geq 1)$$

where $A_n = 1/n \sum_{j=0}^{n-1} V^j$. It follows by LEMMA 9 that $\hat{V}$ and hence V^n, $n \geq 0$, does not increase the L_2 norm of g_k; hence

$$\|Vg_k^* - g_k^*\|_2 \leq \|V(g_k^* - A_n g_k)\|_2 + \|VA_n g_k - g_k^*\|_2 \leq$$

$$\leq \|g_k^* - A_n g_k\|_2 + \|1/n(V^n g_k - g_k) + A_n g_k - g_k^*\|_2 \leq$$

$$\leq 2\|g_k^* - A_n g_k\|_2 + 2/n\|g_k\|_2 .$$

Letting $n \to \infty$ it follows by (4) that $\|Vg_k^* - g_k^*\|_2 = 0$ or

(5) $$Vg_k^* = g_k^* \quad \text{a.e.,} \quad k = 1,2 \ldots$$

Applying (2), (3), (5) and the fact that V does not increase the L_1-norm, we obtain

$$\|Vf^* - f^*\|_1 \leq \|V(f^* - g_k^*)\|_1 + \|Vg_k^* - f^*\|_1 \leq$$

$$\leq \|f^* - g_k^*\|_1 + \|g_k^* - f^*\|_1 =$$

$$= 2\|(f - g_k)^*\|_1 \leq 2\|f - g_k\|_1 \longrightarrow 0$$

as $k \to \infty$; whereupon $Vf^* = f^*$ a.e. □

Note: It follows from inequality (2) in the preceding proof that the equation

$$Pf = f^* , \quad f \in L_1$$

defines a continuous projection of L_1 onto the set of fixed points of V.

COROLLARY. *If in addition to the hypothesis of the* DUNFORD-SCHWARTZ THEOREM $\mu(X) < \infty$, *then for each* $f \in L_1$ *we have in addition to the conclusion (and using the same notation) of that theorem*

$$\|f^* - 1/n \sum_{k=0}^{n-1} V^k f\|_1 \longrightarrow 0 .$$

<u>Proof</u>. The condition $\mu(X) < \infty$ implies $L_\infty \subset L_1$. Write $A_n = 1/n \sum_{k=0}^{n-1} V^k$ and let $g \in L_\infty$. Then by the hypothesis on V A_n does not increase the essential supremum of g, so that

$$|A_n g| \leq \|A_n g\|_\infty \leq \|g\|_\infty < \infty \quad \text{a.e.}$$

Since the function which is identically equal to the constant $\|g\|_\infty$ is integrable and since the sequence $\{A_n g\}$ by the previous theorem is a.e. convergent to a function g^* in L_1, it follows by the Lebesgue dominated convergence theorem that

$$\|A_n g - g^*\|_1 \longrightarrow 0 \ . \tag{1}$$

Now let f be arbitrary in L_1, and using the fact that L_∞ is dense in L_1, choose a sequence $\{g_k\}$ in L_∞ such that

$$\|f - g_k\|_1 \longrightarrow 0 \ .$$

If $f^* = \lim A_n f$ and $g_k = \lim A_n g_k$ as given by the DUNFORD-SCHWARTZ THEOREM, then

$$\|g_k^* - f^*\|_1 = \|(g_k - f)^*\|_1 \leq \|g_k - f\|_1 \ . \tag{2}$$

Furthermore, by the hypothesis on V it follows that

$$\|A_n(f - g_k)\|_1 \leq \|f - g_k\|_1 \ . \tag{3}$$

Therefore

$$\|A_n f - f^*\|_1 \leq \|A_n(f - g_k)\|_1 + \|A_n g_k - g_k^*\|_1 + \|g_k^* - f^*\|_1$$

$$< 2\|f - g_k\|_1 + \|A_n g_k - g_k^*\|_1 \ .$$

Applying (1) with $g = g_k$ we then obtain

$$\overline{\lim} \|A_n f - f^*\|_1 \le 2\|f - g_k\|_1 \longrightarrow 0 \quad \text{as } k \to \infty$$

which implies $\lim \|A_n f - f^*\|_1 = 0$. ▯

The following example shows that mean convergence (of order 1) need not occur if $\mu(X) = \infty$. Let $X = (-\infty,\infty)$, $\mathfrak{A}$ the Lebesgue measurable subsets of X, and μ Lebesgue measure. Define V by

$$(Vf)(x) = f(x - 1) , \quad -\infty < x < \infty ,$$

for every measurable function f. It is easily verified that V satisfies the hypothesis of the DUNFORD-SCHWARTZ THEOREM. In fact $\|Vf\|_p = \|f\|_p$ for $1 \le p \le \infty$ and f measurable, so that V satisfies the hypothesis of LEMMA 6 (LEMMA 9 is not needed here).

Let $f = I_{(0,1]}$. Then $(V^k f)(x) = f(x - k) = I_{(k,k+1]}(x)$, and we obtain

$$1/n \sum_{k=0}^{n-1} V^k f(x) = 1/n \; I_{(0,n]}(x) \longrightarrow 0$$

as $n \to \infty$ for all x. But

$$\|1/n \; I_{(0,n]}\|_1 = 1/n \; \mu(0,n] = 1 \not\longrightarrow 0$$

as $n \to \infty$. Note, however, that if $p > 1$, then

$$\|1/n \; I_{(0,n]}\|_p = \frac{1}{n} \cdot n^{\frac{1}{p}} = \frac{1}{n^{(p-1)/p}} \longrightarrow 0$$

as is predicted by LEMMA 6.

The following remarks are pertinent

(1) Using the same hypothesis as in the DUNFORD-SCHWARTZ THEOREM more can be proved. Let $A_n = 1/n \sum_{k=0}^{n-1} V^k$ (where V denotes both the operator on L_1 and its extension to Λ as given by LEMMA 9), and let f be in some L_p where $1 < p < \infty$; then

$$\bar{f} = \sup_{n\geq 1} |A_n f| \in L_p$$

(c.f. LEMMA 4), and in fact

$$\|\bar{f}\|_p \leq 2\left(\frac{p}{p-1}\right)^{\frac{1}{p}} \|f\|_p \ .$$

A similar result holds when $f \in L_1$ and $\mu(X) < \infty$ (the upper bound for $\|\bar{f}\|_1$, however, is complicated and in practice rather poor). These results imply of course the existence of dominating functions in L_p for the averages $1/n \sum_{k=0}^{n-1} V^k f$. For a proof of these assertions and for other extensions of the theory cf. DUNFORD AND SCHWARTZ [2] pp. 678-708.

(2) The method of proof embodied in LEMMA 3 was first developed jointly by R. V. Chacon and D. Ornstein (c f. CHACON AND ORNSTEIN) who proved the following theorem:

CHACON-ORNSTEIN POINTWISE ERGODIC THEOREM (1960). *Let* V *be a positive linear operator on* L_1 *(i.e.* $Vf \geq 0$ *a.e. whenever* $f \geq 0$ *a.e.), and suppose* $\|Vf\|_1 \leq \|f\|_1$ *for every* $f \in L_1$ *. Then the limit*

$$\lim_n \frac{\sum_{k=0}^{n-1} V^k f}{\sum_{k=0}^{n-1} V^k g}$$

exists a.e. on $[\sum_{k=0}^{\infty} V^k g > 0]$ *for every* f *and* g *in* L_1 , $g \geq 0$ a.e.

Neither the DUNFORD-SCHWARTZ THEOREM nor the CHACON-ORNSTEIN THEOREM implies the other, although recently (1963) by using a considerably refined technique similar to the construction in the proof in this chapter and in the proof of the CHACON-ORNSTEIN THEOREM Chacon has apparently proved a theorem which contains both of the results previously alluded to as special cases (cf. CHACON [2]).

(3) Operators satisfying the hypothesis of either the DUNFORD-SCHWARTZ or the CHACON-ORNSTEIN THEOREMS arise naturally in the theory of Markov processes (in fact this theory was the original motivation for the study, begun by Doob, E. Hopf and others, of operator theoretic generalizations of the classical ergodic theorems). The operator here is defined directly on the function space L_1 by means of an integral equation involving a Markov kernel, e.g.

$$(Vf)(x) = \int f(y)p(x;dy)$$

where p(x;A) is a probability transition mechanism. We shall not go into this, however. (In this connection cf. YOSIDA, p. 379, or DUNFORD AND SCHWARTZ [2], p. 715.)

(4) Needless to say these ergodic theorems are generalizations of the classical ergodic theorems of Birkhoff and von Neumann in which the operators are defined in terms of a transformation acting directly on the ground space X. It will be the purpose of Chapter IV to discuss some aspects of such transformations.

CHAPTER IV

MEASURE PRESERVING TRANSFORMATIONS

Linear operators satisfying conditions of the convergence theorems of the last chapters have classically been those which are induced by transformations acting directly on the underlying ground space X. The purpose of this chapter is to present the basic facts of the theory of measure preserving transformations as it relates to these convergence theorems. In particular the classical ergodic theorem of Birkhoff is proved. We conclude this chapter with a discussion of an example of an ergodic measure preserving transformation arising in the theory of continued fractions.

Most of the material in this chapter is well known. In fact, the example referred to above is given as an exercise in LOÈVE, p. 454. Primarily the material has been arranged and selected towards an original solution to the exercise.

Until otherwise indicated $(X,\mathfrak{A},\mu)$, as in Chapter III, is a σ-finite complete measure space. Let T be a transformation on X with range in X, and suppose T is measurable (i.e. $T^{-1}A \in \mathfrak{A}$ for every $A \in \mathfrak{A}$ where, as usual, $T^{-1}A = \{x : Tx \in A\}$). T is said to be μ-*measure preserving* (or simply *measure preserving*) if

$$\mu(T^{-1}A) = \mu(A) \quad \text{for every } A \in \mathfrak{A} .$$

We shall use the abbreviation m.p. for measure preserving.

Any measurable transformation T on X into X, m.p. or not, induces a transformation V_T on M (the space of complex measurable functions defined a.e. on X) as follows: Letting $f \varepsilon M$, then for any $x \varepsilon X$

$$(V_T f)(x) = f(Tx)$$

provided the right-hand side of this equation is defined. The next lemma is central to the ergodic convergence theorems for m.p. transformations.

LEMMA 1. *Let* T *be m.p. on* X, *and let* V_T *be the induced transformation on* M. *Then* V_T *is linear and positive (i.e.* $f \geq 0$ *a.e. implies that* $V_T f \geq 0$ *a.e.). Moreover,*

(i) $$\int V_T f d\mu = \int f d\mu \quad (f \varepsilon L_1); \quad \text{and}$$

(ii) $$\| V_T f \|_p = \| f \|_p \quad (f \varepsilon L_p, \quad 1 \leq p \leq \infty)$$

that is, V_T *is a linear isometry on each* L_p.

Proof. That V_T is linear and positive is clear from its definition. To prove (i) suppose first that f is an integrable simple function, say $f = \Sigma\, C_k I_{A_k}$; then

(1) $$(V_T f)(x) = \Sigma\, C_k I_{A_k}(Tx) = \Sigma\, C_k I_{T^{-1} A_k}(x)$$

so that

$$(2) \qquad \int V_T f d\mu = \Sigma\ C_k \mu(T^{-1}A_k)$$

$$= \Sigma\ C_k \mu(A_k) = \int f d\mu\ .$$

Now let f be non-negative on X and integrable. We may choose a sequence $\{f_n\}$ of non-negative integrable simple functions such that $f_n \leq f$ and $f_n \to f$ a.e. It follows from (1) and (2) above that $\{V_T fn\}$ is a sequence of non-negative integrable functions. Moreover $\int V_T f_n d\mu = \int f_n d\mu$, and

$$(V_T f_n)(x) = f_n(Tx) \uparrow f(Tx) = (V_T f)(x) \quad \text{a.e.}$$

Applying the monotone convergence theorem it follows that

$$(3) \qquad \int f d\mu = \lim \int f_n d\mu = \lim \int V_T f_n d\mu = \int V_T f d\mu\ .$$

The truth of (i) for an arbitrary $f \in L_1$ may now be seen by writing $f = f_1 - f_2 + i\ (f_3 - f_4)$ where $f_j \geq 0$ a.e., $f_j \in L_1$ $(j = 1,2,3,4)$ and applying (3) to each f_j.

To prove (ii) we consider two cases.

(I) Assume $f \in L_p$ for some $p \in [1, \infty)$. Then for $x \in X$

$$|(V_T f)(x)|^p = |f(Tx)|^p = (V_T |f|^p)(x)\ ,$$

whereupon by (i) $(|f|^p \in L_1)$ we have

$$\|V_T f\|_p^p = \int |V_T f|^p d\mu = \int V_T |f|^p d\mu = \int |f|^p d\mu = \|f\|_p^p\ ,$$

and hence $\|V_T f\|_p = \|f\|_p$.

(II) Assume $f \in L_\infty$. Then for any $a > 0$

$$\mu[|V_T f| \geq a] = \mu(T^{-1}[|f| \geq a]) = \mu[|f| \geq a] ;$$

and therefore

$$\|V_T f\|_\infty = \inf\{a : \mu[|V_T f| \geq a] = 0\}$$

$$= \inf\{a : \mu[|f| \geq a] = 0\} = \|f\|_\infty . \quad \square$$

From LEMMA 1 and the theorems of Chapters II and III we immediately obtain all the usual classical ergodic convergence theorems for m.p. transformations. For example, since L_p is relexive for $1 < p < \infty$, it follows from LEMMA 1 and the KAKUTANI-YOSIDA THEOREM that the averages $1/n \sum_{k=0}^{n-1} V_T{}^k f$ are mean convergent of order p to some function f^* (depending on f) whenever $f \in L_p$ for some $p \in (1,\infty)$, i.e.,

$$\lim_n \int |f^*(x) - 1/n \sum_{k=0}^{n-1} f(T^k x)|^p dx = 0$$

where $dx \equiv d\mu(x) \equiv \mu(dx)$. For $p = 2$ this is just the assertion of the von Neumann mean ergodic theorem proved in 1932. Rather than discuss the formal details of this theorem we shall take up the more interesting case of pointwise convergence.

BIRKHOFF POINTWISE ERGODIC THEOREM (1931). *Let* T *be m.p. on* X. *Then for each* f *in* L_1 *the limit*

$$\lim_n 1/n \sum_{k=0}^{n-1} f(T^k x) = f^*(x)$$

exists for almost all x. *The limit* f^* *is integrable and* $f^*(Tx) = f^*(x)$ *a.e. If* $\mu(X) < \infty$, *then* $\int f^* d\mu = \int f d\mu$.

Proof. Let $f \in L_1$, then

$$\sum_{k=0}^{n-1} f(T^k x) = \sum_{k=0}^{n-1} (V_T{}^k f)(x) \tag{1}$$

where V_T is the operator induced by T. Condition (ii) of LEMMA 1 implies that V_T satisfies the hypothesis of the DUNFORD-SCHWARTZ THEOREM, and hence by (1) the limit of $1/n \sum_{k=0}^{n-1} f(T^k x)$ exists a.e., the limit function f^* is integrable and is a fixed point of V_T, i.e., $f^*(Tx) = V_T f^*(x) = f^*(x)$ a.e.

It remains to show that if $\mu(X) < \infty$, then

$$\int f^* d\mu = \int f d\mu \quad .$$

But the assumption $\mu(X) < \infty$ implies, by the COROLLARY to the DUNFORD-SCHWARTZ THEOREM, that

$$\lim \int 1/n \sum_{k=0}^{n-1} V_T{}^k f d\mu = \int f^* d\mu \ . \tag{4}$$

But by (i) of LEMMA 1 $\int V_T f d\mu = \int f d\mu$, which implies

$$\int 1/n \sum_{k=0}^{n-1} V_T{}^k f d\mu = \int f d\mu$$

for every n = 1,2,..., and therefore by (4)

$$\int f^* d\mu = \lim \int 1/n \sum_{k=0}^{n-1} V_T{}^k f d\mu = \int f d\mu \ . \quad \square$$

Notes. (1) By LEMMA 1 V_T is a positive linear operator on L_1, and $\|V_T f\|_1 = \|f\|_1$ for every $f \in L_1$. It follows that in the case where $\mu(X) < \infty$ we can also obtain the a.e. convergence of $\{1/n \sum_{k=0}^{n-1} V_T{}^k f\}$ ($f \in L_1$) by taking $g \equiv 1$ in the CHACON-ORNSTEIN THEOREM, mentioned in Chapter III; for then $\sum_{k=0}^{n-1} g(T^k x) = n$, so that

$$1/n \sum_{k=0}^{n-1} V_T{}^k f = \frac{\sum_{k=0}^{n-1} V_T{}^k f}{\sum_{k=0}^{n-1} V_T{}^k g}$$

which is a.e. convergent by that theorem.

(2) For an elegant direct proof of the BIRKHOFF THEOREM cf. HALMOS [2], p. 18.

We introduce now the concept of an ergodic transformation. A set $A \in \mathfrak{A}$ is said to be *trivial* (or μ-*trivial*) if either $\mu(A) = 0$ or $\mu(A^c) = 0$. Note that if $\nu \ll \mu$, then a μ-trivial set is also ν-trivial. Let T be a m.p. transformation on X. A function $f \in M$ is *invariant* (with respect to T) if $f(Tx) = f(x)$ a.e., or equivalently, if f is a fixed point of the induced operator V_T; a set $A \in \mathfrak{A}$ is *invariant* (under T) if its indicator I_A is invariant. Clearly a set A is invariant if and only if $\mu(A \Delta T^{-1} A) = 0$. T is *ergodic* if the only invariant sets are trivial. Other terms for ergodic are *indecomposable* or *metrically transitive.*

The following lemma is used in the proof of a corollary to the BIRKHOFF THEOREM. The corollary interprets some of the consequences of ergodicity.

LEMMA 2. *Let* T *be m.p. on* X. *Then* T *is ergodic if and only if every invariant measurable function is constant a.e.*

Proof. Suppose that every invariant measurable function is constant a.e. If $A \in \mathfrak{A}$ is an invariant set, then I_A is an invariant measurable function, and hence either $I_A = 0$ a.e. or else $I_A = 1$ a.e., which implies either $\mu(A) = 0$ or $\mu(A^c) = 0$. Thus every invariant set is trivial, and T is ergodic.

Conversely, suppose that T is ergodic, and let f be an invariant measurable function. Assume first that f is real valued. Since f is finite a.e. on X, it follows that there are real numbers a and b with $a < b$ such that

(1) $$\mu[a \leq f \leq b] \neq 0 .$$

Since f is invariant it follows that the set $[a \leq f \leq b]$ is invariant and hence trivial, since T is ergodic. In view of (1) this implies

(2) $$\mu[f < a] = \mu[f > b] = 0 .$$

Let $\alpha_o = \inf\{\alpha : \mu[f \geq \alpha] = 0\}$.

Then by (1) and (2) it follows that $a \leq \alpha_o \leq b$. Assume $f \neq \alpha_o$ a.e. Then there is an $\varepsilon > 0$ such that either $\mu[f \leq \alpha_o - \varepsilon] \neq 0$ or else $\mu[f \geq \alpha_o + \varepsilon] \neq 0$. The second case cannot occur since $[f \geq \alpha_o + \varepsilon] \subset [f \geq \alpha_o + \frac{1}{2}\varepsilon]$ and by definition of α_o $\mu[f \geq \alpha_o + \frac{1}{2}\varepsilon] = 0$. The first case cannot occur since the invariance of f implies $[f \leq \alpha_o - \varepsilon]$ is trivial, and hence $\mu[f > \alpha_o - \varepsilon] = 0$, again contradicting the definition of α_o. It follows, then, that $f = \alpha_o$ a.e.

Now suppose f is complex valued, measurable, and invariant. Then Re(f) and Im(f) are real and evidently invariant and hence must be constants a.e. by the previous paragraph, whereupon $f = \mathrm{Re}(f) + i\mathrm{Im}(f)$ is constant a.e. ▯

COROLLARY. Let T be an ergodic μ-m.p. transformation on X. If $\mu(X) = \infty$, then

$$\text{(i)} \qquad \lim_n 1/n \sum_{k=0}^{n-1} f(T^k x) = 0 \quad \text{a.e.} \quad (f \in L_1)$$

If $\mu(X) < \infty$, then

$$\text{(ii)} \qquad \lim_n 1/n \sum_{k=0}^{n-1} f(T^k x) = \frac{1}{\mu(X)} \quad f d\mu$$

a.e. for every f in L_1.

<u>Proof</u>. Let $f \in L_1$ and $f^*(x) = \lim_n 1/n \sum_{k=0}^{\infty} f(T^k x)$ which exists a.e. on account of the BIRKHOFF THEOREM. Then f^* is invariant and integrable. Since T is ergodic, it follows by LEMMA 2 that f^* is an integrable constant c.

If $\mu(X) = \infty$, then the only integrable constant function is zero, so that $f^* = c = 0$ a.e. This proves (i).

If $\mu(X) < \infty$, then $c\mu(X) = \int f^* d\mu = \int f d\mu$, and thus

$$c = f^* = \frac{1}{\mu(X)} \int f d\mu \quad \text{a.e.}$$

This proves (ii). $\square$

Note: Conclusion (ii) is usually given the following interpretation: when T is ergodic, one may calculate the "phase average" $\frac{1}{\mu(X)} \int f d\mu$ by calculating arbitrarily large "time averages" $1/n \sum_{k=0}^{n-1} f(T^k x)$ (provided, of course, the x used in the time average is not in the exceptional set of measure zero where everything goes wrong).

EXAMPLE. *An application of the* BIRKHOFF THEOREM *to the theory of continued fractions.*

This example shows that, in general, proving a given transformation T ergodic is usually almost as hard as proving the ergodic theorem itself (cf. the remarks in HALMOS [2], p. 96, in this connection).

In the example $X = [0,1]$, is the algebra of Lebesgue measurable subsets of X, and λ is Lebesgue measure. The transformation in question is defined by

$$Tx = \begin{cases} \frac{1}{x} - [\frac{1}{x}] \text{ , if } x \neq 0 \\ 0, \text{ if } x = 0 \text{ ,} \end{cases}$$

where $[\cdot]$ denotes the greatest integer function. Let μ be the measure on $\mathfrak{A}$ defined by

$$(2) \qquad \mu(A) = \frac{1}{\log 2} \int_A \frac{1}{x+1}\, dx \qquad (A \in \mathfrak{A})$$

($dx = d\lambda(x)$). The measure μ is called *Gauss' measure.* The main theorem, which follows, will be used to derive various results in the measure theory of continued fractions.

THEOREM (I). *T is μ-measure preserving and ergodic.*

Proof that T is μ-m.p. Define μ_T on $\mathfrak{A}$ by

$$\mu_T(A) = \mu(T^{-1}(A)) = \frac{1}{\log 2} \int_{T^{-1}(A)} \frac{1}{x+1}\, dx \qquad (A \in \mathfrak{A})$$

Then μ_T is a measure on $\mathfrak{A}$ (set operations are preserved under T^{-1}). Since any measure on $\mathfrak{A}$ is essentially determined by its values on the semiring of half-open intervals $[a,b)$, $0 \le a < b \le 1$, (cf. HALMOS [1], pp. 30-66), it is clear that in order to prove T is μ-measure preserving it suffices to show

$$(3) \qquad \mu_T[0,t)) = \mu(T^{-1}[0,t)) = \mu([0,t))$$

for every t, $0 < t \le 1$. But for any $t \in (0, 1]$

$$T^{-1}[0,t) = \{x : \frac{1}{x} - [\frac{1}{x}] < t\} =$$

$$= \bigcup_{k=1}^{\infty} \{x : \frac{1}{x} < t + k,\ [\frac{1}{x}] = k\} = \bigcup_{k=1}^{\infty} (\frac{1}{k+t}, \frac{1}{k}] ,$$

and therefore

$$\mu(T^{-1}[0t)) = \frac{1}{\log 2} \sum_{k=1}^{\infty} \int_{\frac{1}{k+t}}^{\frac{1}{k}} \frac{1}{x+1}\,dx = \frac{1}{\log 2} \sum_{k=1}^{\infty} \log \frac{(k+1)(k+t)}{(k+1+t)(k)}$$

$$= \frac{1}{\log 2} \lim_n \sum_{k=1}^{n-1} [\log(\frac{k+1}{k+1+t}) - \log(\frac{k}{k+t})] =$$

$$= \frac{1}{\log 2} \lim_n [\log(\frac{n}{n+t}) - \log(\frac{1}{t+1})] =$$

$$= \frac{1}{\log 2} \log(t+1) = \frac{1}{\log 2} \int_o^t \frac{1}{x+1}\,dx = \mu[0,t) \quad ,$$

which proves (3).

Before proving that T is ergodic we need some preliminary facts from the theory of continued fractions.

An expression of the form

$$(4) \qquad \frac{1}{c_1 + \frac{1}{c_2 + \dots + \frac{1}{c_n}}} = \frac{1|}{|c_1} + \dots + \frac{1|}{|c_n} \ ,$$

where $c_1,\dots,c_n$ are non-zero numbers, is called a *finite continued fraction of order* n with *elements* or *partial quotients* $c_1,\dots,c_n$. Let $\{c_n\}$ be an infinite sequence of non-zero numbers. If the limit of $\frac{1|}{|c_1} + \dots + \frac{1|}{|c_1}$ exists, then it is denoted by $\frac{1|}{|c_1} + \frac{1|}{|c_2} + \dots$ and is called an *infinite continued fraction* with elements $c_1,c_2,\dots$. Note that for any infinite continued fraction we have

$$\frac{1|}{|c_1} + \frac{1|}{|c_2} + \dots = \lim_n\left(\frac{1|}{|c_1} + \dots + \frac{1|}{|c_n}\right) =$$

$$= \frac{1|}{|c_1} + \frac{1|}{|c_2} + \dots + \frac{1|}{\left|c_k + \lim_n\left(\frac{1|}{|c_{k+1}} + \dots + \frac{1|}{|c_{k+n}}\right)\right.}$$

$$= \frac{1|}{|c_1} + \frac{1|}{|c_2} + \dots + \frac{1|}{\left|c_k + \frac{1|}{|c_{k+1}} + \frac{1|}{|c_{k+2}} + \dots\right.}$$

for every $k = 1,2,\dots$ This may be proved by induction using (4). We shall be primarily concerned with simple continued fractions, that is, fractions all of whose elements are positive integers. It is clear from (4) that the value of a simple continued fraction, infinite or not, always lies between 0 and 1.

LEMMA A. *Let* $\{c_n\}$ *be an infinite sequence of positive numbers. Then for every* $x \geq 0$ *and each* $n \geq 2$

$$\frac{1|}{|c_1} + \dots + \frac{1|}{|c_n + x} = \frac{p_{n-1}x + p_n}{q_{n-1}x + q_n} \tag{5}$$

where $\{p_n\}$ and $\{q_n\}$ *do not depend on* x *and are defined by*

$$\text{(6)} \quad \begin{cases} p_1 = 1, \quad p_2 = c_2, \quad p_{k+2} = c_{k+2}\, p_{k+1} + p_k \\ q_1 = c_1, \quad q_2 = c_1 c_2 + 1, \quad q_{k+2} = c_{k+2} q_{k+1} + q_k, \quad k = 1,2\dots \end{cases}$$

The sequences $\{p_n\}$ *and* $\{q_n\}$ *satisfy*

(7) $$p_{n-1}q_n - p_nq_{n-1} = (-1)^n \ , \quad n \geq 2$$

(8) $$p_{n-2}q_n - p_nq_{n-2} = (-1)^{n-1}c_n \ , \quad n \geq 3$$

Proof. For $n = 2$ we have

$$\frac{1 \rvert}{\lvert c_1} + \frac{1 \rvert}{\lvert c_2 + x} = \frac{1}{c_1 + \frac{1}{c_2 + x}} =$$

$$= \frac{x + c_2}{c_1x + c_1c_2 + 1} = \frac{p_1x + p_2}{q_1x + q_2}$$

so that (5) holds for $n = 2$. Assume (5) holds for some $n \geq 2$ and all $x \geq 0$. Then (5) is valid with x replaced by $\frac{1}{c_{n+1} + x}$; and hence

$$\frac{1 \rvert}{\lvert c_1} + \ldots + \frac{1 \rvert}{\lvert c_n} + \frac{1 \rvert}{\lvert c_{n+1} + x} = \frac{1 \rvert}{\lvert c_1} + \ldots + \frac{1 \rvert}{\lvert c_n + \frac{1}{c_{n+1} + x}}$$

$$= \frac{p_{n-1}(\frac{1}{c_{n+1} + x}) + p_n}{q_{n-1}(\frac{1}{c_{n+1} + x}) + q_n} = \frac{p_nx + (c_{n+1}\,p_n + p_{n-1})}{q_nx + (c_{n+1}\,q_n + q_{n-1})} \ ,$$

whereupon by the recursions in (6)

$$\frac{1 \rvert}{\lvert c_1} + \ldots + \frac{1 \rvert}{\lvert c_{n+1} + x} = \frac{p_nx + p_{n+1}}{q_nx + q_{n+1}}$$

and thus (5) holds for $n + 1$ and hence by induction for all $n \geq 2$.

The proofs of (7) and (8) are straightforward induction arguments using (6). We omit the details (cf. HARDY AND WRIGHT, p. 131). ∎

Taking $x = 0$ in (5) we obtain for $n \geq 2$

$$\frac{1|}{|c_1} + \cdots + \frac{1|}{|c_n} = \frac{p_n}{q_n} .$$

(This equation also obviously holds for $n = 1$). The number $\frac{p_n}{q_n}$ is called the *nth order convergent* to the continued fraction $\frac{1|}{|c_1} + \frac{1|}{|c_2} + \cdots$ (whether it exists or not). If $\{c_n\}$ is a sequence of positive integers, then p_n and q_n, $n \geq 1$, are also positive integers which, by induction and using (7), are relatively prime; and, therefore, the fraction $\frac{p_n}{q_n}$ is in reduced form.

LEMMA B. *Let* $\{c_n\}$ *be any infinite sequence of positive integers, and let* $\{p_n\}$ *and* $\{q_n\}$ *be the sequences of positive integers defined as in* (6). *Then the even convergents* $\{\frac{p_{2n}}{q_{2n}}\}$ *are strictly increasing, the odd convergents* $\{\frac{p_{2n-1}}{q_{2n-1}}\}$ *are strictly decreasing, and every even convergent is less than every odd convergent. Furthermore the limit as* $n \to \infty$ *of*

$$\frac{p_n}{q_n} = \frac{1|}{|c_1} + \cdots + \frac{1|}{|c_n}$$

exists, and, letting x_o *denote this limit, we have*

$$\frac{p_{2n}}{q_{2n}} < x_o < \frac{p_{2n-1} + p_{2n}}{q_{2n-1} + q_{2n}} < \frac{p_{2n-1}}{q_{2n-1}} \tag{9}$$

for $n = 1,2,\ldots$.

Proof. Taking n = 2k in (8), we have

$$\frac{p_{2k-2}}{q_{2k-2}} - \frac{p_{2k}}{q_{2k}} = \frac{(-1)^{2k-1}c_{2k}}{q_{2k-2}\,q_{2k}} < 0 ,$$

so that $\{\frac{p_{2n}}{q_{2n}}\}$ is a strictly increasing sequence. Similarly $\{\frac{p_{2n-1}}{q_{2n-1}}\}$ is a strictly decreasing sequence. Taking n = 2k in (7), we obtain

$$(10) \qquad \frac{p_{2k-1}}{q_{2k-1}} - \frac{p_{2k}}{q_{2k}} = \frac{1}{q_{2k}\,q_{2k-1}} > 0 .$$

This, together with the preceding remarks, implies that every even convergent is less than every odd convergent.

Now, by (6), $q_1 = c_1 \geq 1$ and $q_{n+1} = c_{n+1}\,q_n + q_{n-1} > q_n$; whereupon by induction $q_n > n$. Letting $x' = \lim_n \frac{p_{2n}}{q_{2n}}$ and $x'' = \lim_n \frac{p_{2n-1}}{q_{2n-1}}$, it follows by (10) that

$$0 \leq x'' - x' = \lim_n\left(\frac{p_{2n-1}}{q_{2n-1}} - \frac{p_{2n}}{q_{2n}}\right) \leq \lim_n \frac{1}{2n(2n-1)} = 0$$

or $x' = x''$, which clearly implies that $\lim_n \frac{p_n}{q_n}$ exists.

To prove (9) let $r_k = \frac{1|}{|c_{k+1}} + \frac{1|}{|c_{k+2}} + \dots$. Then it follows that $0 < r_k < 1$ and

$$(11) \qquad x_o = \frac{1|}{|c_1} + \frac{1|}{|c_2} + \dots + \frac{1|}{|c_k + r_k} = \frac{p_{k-1}r_k + p_k}{q_{k-1}r_k + q_k}$$

where the last equality follows from (5) of LEMMA A. Now define a function g on $[0,\infty)$ by

$$g(x) = \frac{p_{2n-1}x + p_{2n}}{q_{2n-1}x + q_{2n}} \quad .$$

Then, applying (7),

$$g'(x) = \frac{p_{2n-1}\, q_{2n} - q_{2n-1}\, p_{2n}}{(q_{2n-1}x + q_{2n})^2} = \frac{1}{(q_{2n-1}x + q_{2n})^2} > 0 \ ,$$

so that g is strictly increasing on $[0,\infty)$. Hence

$$g(0) < g(r_{2n}) < g(1) < g(\infty-) \ .$$

But $g(0) = \dfrac{p_{2n}}{q_{2n}}$, by (11) $g(r_{2n}) = x_o$, $g(1) = \dfrac{p_{2n-1} + p_{2n}}{q_{2n-1} + q_{2n}}$ and $g(\infty-) = \dfrac{p_{2n-1}}{q_{2n-1}}$. □

Note. From (11) in the preceding proof and by (7) it follows that

$$\left| x_o - \frac{p_m}{q_m} \right| = \left| \frac{p_m\, r_{m+1} + p_{m+1}}{q_m\, r_{m+1} + q_{m+1}} - \frac{p_m}{q_m} \right|$$

$$= \left| \frac{p_m\, q_{m+1} - p_{m+1}\, q_m}{q_m(q_m\, r_{m+1} + q_{m+1})} \right| < \frac{1}{q_m\, q_{m+1}}$$

where, as before, $r_k = \dfrac{1|}{|c_{k+1}} + \dfrac{1|}{|c_{k+2}} + \dots$. This inequality is the basic inequality in the theory of approximation by continued fractions.

The following theorem, which we include primarily for completeness, and the previous lemmas are at the basis of any study of simple continued fractions.

THEOREM (II). *To every irrational number* x, $0 < x < 1$, *there is a unique simple continued fraction with value equal to* x. *To every rational number* x, $0 < x < 1$, *there is a unique finite simple continued fraction with value equal to* x *in which the last element is greater than one.*

We shall not go into the proof per se but rather we give the algorithm for generating the elements $\{c_n\}$ for the fraction corresponding to a given irrational number x_o, deriving in the process the necessary facts needed to prove T is ergodic.

For a complete proof and for more detailed information concerning simple continued fractions cf. HARDY AND WRIGHT, pp. 129-153 or KHINCHIN.

Recall from (1) that the transformation T was defined by

$$Tx = \begin{cases} \frac{1}{x} - [\frac{1}{x}] , & x \neq 0 \\ 0 , & x = 0 . \end{cases}$$

We define positive interger-valued functions $\{c_n\}$ on $[0,1]$ as follows

$$(12) \qquad \begin{cases} c_1(x) = [\frac{1}{x}] , \quad 0 < x \leq 1, \quad c_1(0) = \infty \\ c_k(x) = c_1(T^{k-1}x) = [\frac{1}{T^{k-1}x}] , \quad k = 1,2... \end{cases}$$

where T^k is the kth iteration of T. (Since in the context x will be irrational, $c_k(0)$ plays no role here.) If x_o is an irrational number with $0 < x_o < 1$, then by induction it follows that $T^k x_o$, $k \geq 0$, is irrational. Hence $\{c_k(x_o)\}$ is an infinite sequence of positive integers, and moreover we have

LEMMA C. *Let* x_o *be irrational,* $0 < x_o < 1$. *Write* $c_n = c_n(x_o)$, $n = 1,2,\ldots$, *and let* $\{\frac{p_n}{q_n}\}$ *be the corresponding sequence of nth order convergents as defined by* (6). *Then* for each $n = 1,2,\ldots$

(13)
$$x_o = \frac{1|}{|c_1} + \ldots + \frac{1|}{|c_n + T^n x_o} = \frac{p_{n-1}\, T^n x_o + p_n}{q_{n-1}\, T^n x_o + q_n} = \frac{1|}{|c_1} + \frac{1|}{|c_2} + \ldots ,$$

(14)
$$T^n x_o = \frac{1|}{|c_{n+1}} + \frac{1|}{|c_{n+2}} + \ldots .$$

(15) *If*
$$y = \frac{p_{n-1}x + p_n}{q_{n-1}x + q_n}$$

for some $n \geq 2$ and $0 < x < 1$, *then* $T^n y = x$.

<u>Proof</u>. By definition of $c_k(x_o)$ we have

$$c_1 = [\frac{1}{x_o}] = \frac{1}{x_o} - Tx_o$$

which implies

$$x_o = \frac{1}{c_1(x_o) + Tx_o} \quad . \tag{16}$$

Replacing x_o in (16) by Tx_o we obtain

$$Tx_o = \frac{1}{c_1(Tx_o) + T^2x_o} = \frac{1}{c_2(x_o) + T^2x_o} \quad .$$

Substituting this in (16), it follows that

$$x_o = \cfrac{1}{c_1(x_o) + \cfrac{1}{c_2(x_o) + T^2x_o}} = \frac{1|}{|c_1} + \frac{1|}{|c_2 + T^2x_o}$$

Continuing in this fashion, we obtain the first equality of (13). The second equality follows from (5) of LEMMA A.

Since the function

$$g(x) = \frac{p_{2n-1}x + p_{2n}}{q_{2n-1}x + q_{2n}}$$

is strictly increasing with x (cf. the proof of LEMMA B) and since $T^{2n}x_o$ is an irrational number strictly between 0 and 1, it follows that

$$\frac{p_{2n}}{q_{2n}} < x_o = \frac{p_{2n-1}\,T^{2n}x_o + p_{2n}}{q_{2n-1}\,T^{2n}x_o + q_{2n}} < \frac{p_{2n-1} + p_{2n}}{q_{2n-1} + q_{2n}} < \frac{p_{2n-1}}{q_{2n-1}}$$

for every $n = 1,2,\dots$. Letting $x_o' = \lim_m \frac{p_m}{q_m}$, which exists by LEMMA B, then

$$x_o' = \lim_n \frac{p_{2n}}{q_{2n}} \le x_o \le \lim_n \frac{p_{2n-1}}{q_{2n-1}} = x_o'$$

which completes the proof of (13).

Equation (14) now follows immediately from (13).

To prove (15) note that if

$$y = \frac{p_{n-1}x + p_n}{q_{n-1}x + q_n} \quad ,$$

then by (7) $y = \frac{1|}{|c_1} + \dots + \frac{1|}{|c_n + x}$, from which it easily follows that $T^n y = x$. ▯

Note. Equation (14) implies that T may be construed as a "shift" transformation on an appropriately chosen sequence space, as most measure preserving ergodic transformations can.

We may now complete the proof of THEOREM (I).

Proof that T is ergodic. Let A be any measurable subset of [0,1] invariant under T. To prove T ergodic we must show that A is μ-trivial where μ is Gauss' measure defined by (2). Since the set of rationals in [0,1] have μ-measure zero, we may assume that A consists entirely of irrational numbers. If $\lambda(A) = 1$, then clearly $\mu(A) = 1$, and there is nothing to prove since $\mu(X) = 1$ and hence $\mu(A^c) = \mu(X) - \mu(A) = 0$. Suppose, then $\lambda(A) = r$. We will show that $r = 0$ and hence $\mu(A) = 0$.

Let x_o be any irrational number with $0 < x_o < 1$, let $n \ge 1$ be an arbitrary integer and let $\frac{p_{2n}}{q_{2n}}$ and $\frac{p_{2n-1}}{q_{2n-1}}$ be the $2n^{th}$ and (2n-1)th convergents to x_o as in LEMMA C. Set $a = \frac{p_{2n}}{q_{2n}}$ and $b = \frac{p_{2n-1} + p_{2n}}{q_{2n-1} + q_{2n}}$ and let

$$g(x) = \frac{p_{2n-1}x + p_{2n}}{q_{2n-1}x + q_{2n}}, \quad 0 \le x \le 1 .$$

Then g is strictly increasing on [0,1] and has range [a,b], and by (15) of LEMMA C it follows that

$$T^{2n}(g(x)) = x , \quad 0 \le x \le 1 .$$

Whereupon, since A is invariant,

$$I_A(g(x)) = I_A(T^{2n}(g(x))) = I_A(x) , \quad 0 \le x \le 1$$

Combining these facts we obtain,

$$\lambda(A \cap [a,b]) = \int_a^b I_A(y)dy = \int_0^1 I_A(g(x))g'(x)dx =$$

$$= \int_0^1 I_A(x) \frac{(p_{2n-1}\, q_{2n} - p_{2n}\, q_{2n-1})}{(q_{2n-1}x + q_{2n})^2} dx =$$

$$= \frac{1}{q_{2n}^2} \int_0^1 I_A(x) \frac{dx}{\left(\frac{q_{2n-1}}{q_{2n}} x + 1\right)^2}$$

Now $\lambda([a,b]) = \dfrac{p_{2n-1} + p_{2n}}{q_{2n-1} + q_{2n}} - \dfrac{p_{2n}}{q_{2n}} = \dfrac{1}{q_{2n}(q_{2n-1} + q_{2n})}$, and, letting $t' = \dfrac{q_{2n-1}}{q_{2n}}$, then $0 < t' < 1$ ($\{q_k\}$ is obviously an increasing sequence by (6)), so that we obtain

$$(17) \qquad \frac{\lambda(A\cap[a,b])}{\lambda([a,b])} = \frac{q_{2n}(q_{2n-1}+q_{2n})}{q_{2n}^2}\int_0^1 I_A(x)\,\frac{dx}{(t'x+1)^2} =$$

$$= (1+t')\int_0^1 I_A(x)\,\frac{dx}{(t'x+1)^2} \leq$$

$$\leq \max_{0\leq t\leq 1}(1+t)\int_0^1 I_A(x)\,\frac{dx}{(tx+1)^2}$$

Consider the continuous function on [0,1]

$$f(t) = (1+t)\int_0^1 I_A(x)\,\frac{dx}{(tx+1)^2} \quad .$$

The function f attains its maximum value say $\varepsilon_o > 0$ at some point $t_o\, \varepsilon[0,1]$. If $t_o = 0$, then

$$\varepsilon_o = f(0) = \int_0^1 I_A(x)dx = \lambda(A) = r < 1 \ .$$

If $0 < t_o \leq 1$, then, since $I_A = 0$ on a set A^c of positive measure, it follows that

$$\varepsilon_o = f(t_o) = (1+t_o)\int_0^1 I_A(x)\,\frac{dx}{(t_ox+1)^2} <$$

$$< (1+t_o)\int_0^1 \frac{dx}{(t_ox+1)^2} = 1$$

In either case we have by (17)

$$\frac{\lambda(A\cap[a,b])}{\lambda([a,b])} \leq \max_{0\leq t\leq 1} f(t) = \varepsilon_o < 1 \tag{18}$$

For each irrational x_o in (0,1) and each integer $n \geq 1$ let

$$\Delta(x_o,n) = [\frac{p_{2n}}{q_{2n}}, \frac{p_{2n-1}+p_{2n}}{q_{2n-1}+q_{2n}}] . \tag{19}$$

Then for fixed x_o it follows by LEMMAS B and C that the intervals $\{\Delta(x_o,n)\}$ are nested and

$$\{x_o\} = \bigcap_{n=1}^{\infty}\Delta(x_o,n)$$

By (18) we also have

$$\lambda(A\cap\Delta(x_o,n)) \leq \varepsilon_o\ \lambda(\Delta(x_o,n)) \tag{20}$$

for every x_o, $n \geq 1$, where $0 < \varepsilon_o < 1$ and ε_o does not depend on x_o or n. Let $\delta > 0$ be arbitrary, and let G be an open set in (0,1) such that

$$A \subset G \quad \text{and} \quad \lambda(G) \leq \lambda(A) + \delta \tag{21}$$

By the preceding remarks it follows that for each x_o in A and for all sufficiently large n $\Delta(x_o,n) \subset G$. The family F of all such intervals clearly covers A in the sense of Vitali (cf. HEWITT, p. 262), and therefore by Vitali's covering theorem there exists a countable collection $\{J_k\}$ of disjoint intervals in F such that

$$\lambda(A - \bigcup_{k=1}^{\infty} J_k) = 0 \ .$$

Now by (20) $\lambda(J_k \quad A) \leq \varepsilon_o \lambda(J_k)$; and, by definition of the family F, $J_k \subset G \quad k = 1,2,\ldots$. Hence

$$\lambda(A) = \lambda(\bigcup_{k=1}^{\infty} AJ_k) = \sum_{k=1}^{\infty} \lambda(AJ_k) \leq$$

$$\leq \varepsilon_o \sum_{k=1}^{\infty} \lambda(J_k) = \varepsilon_o \lambda(\bigcup_{k=1}^{\infty} J_k) \leq$$

$$\leq \varepsilon_o \lambda(G) \leq \varepsilon_o (\lambda(A) + \delta) \ ,$$

where the last inequality is from (21). Since $\delta > 0$ is arbitrary, it follows that

$$\lambda(A) \leq \varepsilon_o \lambda(A) < \lambda(A) \ ,$$

since $\varepsilon_o < 1$ unless $\lambda(A) = 0$. Hence $\lambda(A) = 0$. □

Applications. Since T is μ-measure preserving and ergodic, it follows by the definition of Gauss' measure μ (cf. (2)) and the COROLLARY to the BIRKHOFF THEOREM that for any measurable function g on [0,1]

$$(22) \qquad 1/n \sum_{k=1}^{n-1} g(T^k x) \xrightarrow{\text{a.e.}} \frac{1}{\log 2} \int_0^1 \frac{g(x)}{x+1}\, dx \ ,$$

provided the integral on the right of (22) exists. Using (22), we can prove the following theorem (cf. KHINCHIN, p. 86) for which a different proof not using the ergodic theory is given.

THEOREM (III). *Suppose that* f *is a function defined on the positive integers and suppose there are constants* M *and* δ *such that*

$$|f(r)| < Mr^{\frac{1}{2}-\delta} \qquad (r = 1,2,\ldots).$$

Then

$$(23) \qquad 1/n \sum_{k=1}^{n} f(c_k(x)) \xrightarrow{\text{a.e.}} \frac{1}{\log 2} \sum_{r=1}^{\infty} f(r)\log[1 + \frac{1}{r(r+2)}] ,$$

where $c_k(x)$ *is the* k^{th} *element in the continued fraction expansion of* x.

<u>Proof</u>. Let $g(x) = f(c_1(x))$. Then $g(x) = f(r)$ when $c_1(x) = [\frac{1}{x}] = r$; that is $g(x) = f(r)$, a constant, on the interval $\frac{1}{r+1} < x \le \frac{1}{r}$. Hence

$$\int_0^1 \frac{g(x)}{x+1}\,dx = \sum_{r=1}^{\infty} f(r) \int_{\frac{1}{r+1}}^{\frac{1}{r}} \frac{1}{x+1}\,dx =$$

$$= \sum_{r=1}^{\infty} f(r) \log[1 + \frac{1}{r(r+2)}] .$$

The condition $|f(r)| < Mr^{\frac{1}{2}-\delta}$ implies the absolute convergence of the last written series. Therefore by (22) and (12)

$$\lim_n 1/n \sum_{k=0}^{n-1} g(T_x^k) = \lim_n 1/n \sum_{k=1}^{n} f(c_k(x)) =$$

$$= \frac{1}{\log 2} \sum_{r=1}^{\infty} f(r)\log[1 + \frac{1}{r(r+2)}] \quad \text{a.e.} \quad \square$$

Taking $f(r) = 1$ if $r = p$ (p a positive integer) and $f(r) = 0$ otherwise, we have by (23)

$$(24) \qquad 1/n \sum_{k=1}^{n} f(c_k(x)) \xrightarrow{\text{a.e.}} \frac{1}{\log 2} \log \frac{(p+1)^2}{p(p+2)}$$

This implies that the relative frequency of occurrence of the integer p (the limiting value of the average on the left in (24)) in the sequence of elements $\{c_n(x)\}$ of the expansion of x has the value $\frac{1}{\log 2} \log \frac{(p+1)^2}{p(p+2)}$ for almost all x in $[0,1]$.

Taking $f(r) = \log r$ in (23), we have

$$1/n \sum_{k=1}^{n} \log c_k(x) \xrightarrow{\text{a.e.}} \frac{1}{\log 2} \sum_{r=1}^{\infty} \log r \log \left[1 + \frac{1}{r(r+2)}\right]$$

which implies

$$(25) \qquad \sqrt[n]{c_1(x)c_2(x)\cdots c_n(x)} \xrightarrow{\text{a.e.}} \prod_{r=1}^{\infty} \left(1 + \frac{1}{r^2+2r}\right)^{\frac{\log r}{\log 2}} \doteq 2.6$$

That is, the geometric means of the elements in the continued fraction expansions for almost all x in $[0,1]$ have limiting value approximately 2.6.

Since $\int_0^1 \frac{c_1(x)}{x + 1} dx = + \infty$, a simple truncation argument shows that

$$1/n \sum_{k=1}^{n} c_k(x) \xrightarrow{\text{a.e.}} + \infty .$$

Thus the arithmetic means of the elements are a.e. divergent (compare this with (25)). This also implies that the elements themselves are almost everywhere unbounded, i.e., $\overline{\lim}\, c_n(x) = +\infty$ a.e.

The truncation argument runs as follows. Let $f \geq 0$ a.e. and $\int f d\mu = +\infty$. If $f_m = f^{m-}$ (cf. Chapter III); then $0 \leq f_m \leq m$ and $f_m \leq f$. Hence $f_m \in L_1$, and we have

$$\underline{\lim}_n\, 1/n \sum_{k=0}^{n-1} f(T^k_x) \geq \lim_n\, 1/n \sum_{k=0}^{n-1} f_m(T^k_x) =$$

$$= \frac{1}{\mu(x)} \int f_m d\mu \quad \text{a.e.}$$

whereupon it follows that

$$1/n \sum_{k=0}^{n-1} f(T^k x) \xrightarrow{\text{a.e.}} +\infty ,$$

since $\lim_m \int f_m d\mu = \int f d\mu = +\infty$.

For some deeper results in the measure theory of continued fractions cf. KHINCHIN, pp. 51-95, or BILLINGSLEY, pp. 40-50. This latter reference also contains an alternative proof of the ergodicity of T.

BIBLIOGRAPHY

BILLINGSLEY, P., *Ergodic Theory and Information,* New York: Wiley, 1965.

CHACON, R. V. [1], "An Ergodic Theorem for Operators Satisfying Norm Conditions," *Journal of Mathematics and Mechanics,* Vol. 11, 1962.

_______ [2], "Convergence of Operator Averages," *Ergodic Theory,* Fred B. Wright, editor, New York: Academic Press, 1963.

CHACON, R. V. and Ornstein, D. S., "A General Ergodic Theorem," *Illinois Journal of Mathematics,* Vol. 4, 1960.

DUNFORD, N. and Schwartz, J. T. [1], "Convergence Almost Everywhere of Operator Averages," *Journal of Rational Mechanics and Analysis,* Vol. 5, 1956.

_______ [2], *Linear Operators,* Vol. I, New York: Interscience, 1958.

HALMOS, P. R. [1], *Measure Theory,* Princeton, N. J.: Van Nostrand, 1950.

_______ [2], *Lectures on Ergodic Theory,* New York: Chelsea, 1956.

_______ [3], "Recent Progress in Ergodic Theory," *Bulletin of the American Mathematical Society,* Vol. 67, 1961.

HARDY, G. H. and Wright, E. M., *An Introduction to the Theory of Numbers,* 4th ed., Oxford: Clarendon Press, 1959.

HEWITT, E. and Stromberg, K., *Real and Abstract Analysis,* New York: Springer-Verlag, 1965.

KAKUTANI, S. and Yosida, K. "Operator-Theoretical Treatment of Markoff Process and Mean Ergodic Theorem," *Annals of Mathematics,* Vol. 42, 1941.

KHINCHIN, Y. A., *Continued Fractions,* Chicago: University of Chicago Press, 1964.

LOÈVE, M., *Probability Theory,* 3rd ed., Princeton, N. J.: Van Nostrand, 1963.

TAYLOR, A. E., *Introduction to Functional Analysis,* New York: Wiley, 1958.

YOSIDA, K., *Functional Analysis,* New York: Academic Press, 1965.

ZYGMUND, A., *Trigonometric Series,* Vol. II, Cambridge: Cambridge University Press, 1959.